UNIVERSITE SAAD DAHLAB DE BLIDA

Faculté des Sciences
Département d'informatique

MEMOIRE DE MAGISTER

Spécialité : Systèmes d'information et de connaissances

DETECTION DE LA PEAU HUMAINE DANS DES IMAGES COULEURS A L'AIDE DE LA REGRESSION LOGISTIQUE BAYESIENNE A NOYAU MULTINOMIALE

Par

Idir FILALI

Devant le jury composé de :

S. Oukid	Maître de conférences A, U. de Blida	Présidente
L. Hamami	Professeur, E.N.P., Alger	Examinatrice
Y. Chibani	Professeur, U.S.T.H.B., Alger	Examinateur
N. Benblidia	Maître de conférences A, U. de Blida	Promotrice
D. Ziou	Professeur, U. de Sherbrooke, Canada	Co-promoteur
F.Z. Reguieg	Maître assistante A, U. de Blida	Invitée

Blida, Janvier 2012

RESUME

La détection de la peau est une étape importante dans de nombreux algorithmes de vision par ordinateur. Il est habituellement un processus qui commence au niveau du pixel et qui implique un pré-processus de transformation colorimétrique suivie par un processus de classification. Il s'agit d'une tâche préliminaire de base utilisée dans de nombreux domaines comme en télésurveillance et en reconnaissance de forme. Nous présentons dans ce mémoire un détecteur de peau basé sur une approche nommée régression logistique bayésienne à noyaux (RLBN). Dans le cadre de cette approche, le noyau de Fisher est exploité pour faire une projection des données d'origines présentées dans l'espace de couleur d'origine dans un nouvel espace de dimension beaucoup plus grande appelé espace de caractéristiques (feature space) de telle manière à pouvoir discriminer les classes considérées (classe peau et classe non peau) linéairement d'une manière plus efficace. La régression logistique sert de modèle de prédiction et s'emploie dans le cadre de la discrimination des pixels peau et ceux de non peau. Le théorème de Bayes est utilisée afin d'estimer les valeurs optimales des paramètres de l'équation logistique. Nous nous sommes intéressés à traiter trois types de couleurs de peau appartenant à différentes races ethniques: la race jaune, la race noire et la race blanche. A cet effet, nous avons construit un classifieur spécifique à chaque race puis nous avons combiné les trois classifieurs en un seul dit classifieur de synthèse dédié à la détection de tout type de peau. Nous avons utilisé la base d'image Compaq pour évaluer notre classifieur et les résultats obtenus sur les performances de détection de notre classifieur montrent son efficacité en comparaison avec ceux obtenus par les classifieurs basés sur des méthodes proposées précédemment et évalués sur la même base d'images que la notre.

REMERCIEMENTS

Mes vifs remerciements s'adressent en premier lieu à mes encadreurs Mr Ziou Djemel, professeur à l'université de Sherbrooke et Mme Benblidia Nadjia maître de conférences à l'université de Blida pour m'avoir proposé ce sujet, dirigé mes travaux, et permis de développer mon sens de la recherche. Qu'ils soient assurés de ma profonde gratitude et de ma haute considération.

Mon profond respect et mes cordiaux remerciements aux membres du jury, Mme Oukid Saliha, Maître de conférences à l'USDB, Mme Hamami Latifa, professeur à l'ENP et Mr Chibani Youcef, professeur à l'USTHB qui me font l'honneur d'examiner et de juger mon travail.

Merci

TABLE DES MATIERES

TABLE DES FIGURES

LISTE DES TABLEAUX

INTRODUCTION

1. Contexte

La vision par ordinateur est un domaine de recherche qui n'a cessé de se développer depuis le début des années 40 et qui trouve aujourd'hui des applications dans de nombreux secteurs d'activité. Les systèmes d'imagerie, caméras et systèmes de vision, sont de plus en plus accessibles et performants, et induisent des progrès considérables dans plusieurs domaines. Des développements importants ont été réalisés, mais la vision par ordinateur reste un champ d'investigation très actif avec de nombreux problèmes difficiles et non entièrement résolus.
D'une manière générale, la vision par ordinateur peut être considérée comme un processus de traitement de l'information qui est issue d'images numérisées [61]. Les questions qui se posent alors concernent la nature de ces informations et leur représentation: quelle sorte d'information extraire de l'image? Comment décrire et/ou représenter cette information pour en faciliter l'interprétation? Il est clair que la nature et la représentation des informations dépendent de l'application envisagée. Cependant, dans tout processus d'analyse d'image, il faut pouvoir extraire certaines parties de l'image, mesurer des propriétés de ces parties ou des relations entre ces parties, et utiliser les valeurs de ces propriétés pour interpréter le contenu de l'image [172]. Ces trois étapes d'extraction, de mesure et d'interprétation, sont présentes dans presque toutes les applications [184].

La détection de la peau est un problème de vision par ordinateur qui consiste à pouvoir localiser et détecter la présence de personnes ou toute partie du corps humain dans des images ou vidéos numérique d'une manière automatique. L'approche de détection la plus utilisée est celle basée sur les caractéristiques colorimétriques des images. Celle-ci agit au niveau pixels. Un pixel correspond à la plus petite unité de surface permettant de mesurer la définition d'une image numérique matricielle.
La détection de la peau humaine par une approche couleur représente un processus préliminaire de base pour beaucoup de champs d'application comme l'identification de personne, la vidéo surveillance et le filtrage des images sur le Web.

Il est en général constitué de deux phases: la phase d'entrainement et la phase de test. L'entrainement du détecteur de peau consiste d'abord à trouver une base d'images adéquate répondant aux objectifs fixés, puis à choisir un espace de couleur de conversion et à effectuer l'apprentissage du système de classification sur celui-ci. La phase de test consiste d'abord à convertir l'image à tester dans le même espace de couleur utilisé dans la phase d'entrainement, puis à effectuer l'application du

système de classification entrainé pour la discrimination des pixels peau de ceux de non peau. Le résultat d'une telle opération est une image segmentée en régions peau et régions non peau. C'est précisément dans ce contexte que s'inscrivent les travaux que nous avons développés dans le cadre de ce mémoire. Nous nous sommes principalement intéressés au problème d'identification de pixels peau dans l'image. Ce sujet de recherche est d'enjeu important dans la mesure où il est indispensable avant d'envisager des analyses et des traitements de niveau supérieur. Notre objectif est de construire un modèle de peau générique qui pourrait servir ensuite de base pour le développement de beaucoup d'applications. La mise en place de tels systèmes doit, au préalable, s'appuyer sur un système de détection et de segmentation des régions de peau.

2. Problématique et objectifs

Dans la littérature, de nombreux travaux ont été réalisés sur l'élaboration d'un modèle de peau basé sur l'indice couleur. La difficulté réside dans la prise en compte des conditions d'illumination variées, de la richesse ethnique avec des teintes variées et des décors complexes. En effet, la plupart des travaux pratiquent une phase d'apprentissage sur des classes prédéfinies d'images, sous des conditions d'éclairage connues à l'avance. Si ces modèles conviennent généralement aux systèmes à base d'images peu variées, ils sont peu adaptés aux systèmes contenant une grande variété d'images.

Nous avons utilisé la base d'image Compaq [117] pour l'évaluation de notre classifieur. Le choix d'une telle base est motivé par ses caractéristiques qui répondent pleinement aux objectifs que nous avons fixés. Le but de notre approche est de pouvoir détecter les pixels peau dans des images numériques caractérisées par l'hétérogénéité des conditions d'acquisition des images comme l'illumination variante, l'acquisition sous différents angles et positions, la diversité des appareils de capture des images avec différentes caractéristiques, la diversité des sources d'acquisition des images ainsi que la diversité ethnique des personnes représentées sur ces images. Cela rend la tâche de classification particulièrement difficile. Plusieurs chercheurs ont eu recours à la base d'images Compaq pour la validation de leurs méthodes [145] [66]. Nous nous sommes intéressé à améliorer la précision de détection des classifieurs de peau basés sur ces méthodes. Nous situeront les résultats de notre approche par rapport à ces dernière et nous effectueront une étude comparative sur le plan des performances de détection.

3. Notre démarche

Nous avons opté pour un système de détection basé sur les travaux de Ksantini et al. [131] pour construire notre système de détection. Il s'agit de la régression logistique bayésienne à noyaux (*RLBN*). Celui-ci est basé sur les noyaux de Fisher afin d'effectuer une projection linéaire de la représentation des données de l'espace colorimétrique d'origine dans un nouvel espace de dimension généralement plus grand appelé espace de *caractéristiques* de manière à pouvoir discriminer les pixels

peau des pixels non peau de façon plus efficace. Parmi les méthodes basées sur les noyaux de Fisher nous citons la régression logistique à noyaux et les SVM (Support Vector Machine). L'approche RLBN se base sur le théorème de Bayes pour une estimation optimale des paramètres de la régression logistique à noyaux utilisée durant la phase d'apprentissage.

Afin d'améliorer les performances de détection de notre classifieur, nous avons défini une stratégie de fusion de plusieurs classifieurs, chacun dédié à la détection d'une race ethnique. Le résultat de détection final sera la combinaison des résultats de détection de ces classifieurs. Nous avons à cet effet opté pour une régression logistique bayésienne à noyaux multinomiale.

4. Organisation du mémoire

L'organisation du présent manuscrit est présentée ci-dessous.

Le chapitre 1 aborde le domaine de la détection de la peau d'une manière générale, recense les différents courants d'approches qui y sont suivies et présente l'apport des espaces de couleurs pour la tâche de détection. Une description de ces espaces est effectuée dans ce chapitre.

Le chapitre 2 est consacré à l'état de l'art sur les travaux d'identification des pixels peau dans les images couleur numérisées. Les différentes méthodes ont été classées et présentées en plusieurs types d'approches. Leurs avantages et défauts y sont discutés.

Le chapitre 3 traite les méthodes d'apprentissage basées sur l'estimation de paramètres durant la phase d'apprentissage les plus utilisées notamment celles qui ont une relation avec notre travail à savoir la régression logistique simple et bayésienne ainsi que les SVM.

Le chapitre 4 est dédié à la définition de notre modèle de peau. Nous présenterons la méthode adoptées avant d'aborder nos études expérimentales permettant de situer les performances de détection de notre approche en comparaison avec celles des autres méthodes dont l'évaluation a été porté sur la même base d'images que la notre.

CHAPITRE 1
LA DETECTION AUTOMATIQUE DE LA PEAU

1.1 Introduction

1.1.1 Présentation générale de la détection de la peau

La détection de la peau est un moyen efficace de localiser une personne dans une image car au moins certaines parties du corps comme la tête, les mains, les bras et le visage sont visibles. Cela revient à détecter les pixels correspondant à une peau humaine dans une image couleur. Cette détection repose sur des techniques de classification. Certaines approches s'intéressent à définir une classe pour les pixels peau et une autre pour les pixels non peau. La mise en œuvre de ce principe donne comme résultat une image binaire de même taille que l'image à traiter avec l'attribution d'une valeur à la peau et une autre pour l'arrière plan. La figure suivante montre une image couleur et son masque binaire après détection des régions de peau.

a b

Figure. 1.1: (a) et (b) représentent respectivement l'image originale et son masque binaire.

D'autres approches s'intéressent à identifier uniquement la classe peau. Le résultat de ce principe est une image de même taille que l'original avec l'attribution d'une valeur colorimétrique à la classe peau uniquement sans changement dans les valeurs des pixels n'appartenant pas à cette classe. La figure 1.2 [17] montre une image couleur et le résultat de la classification après détection des régions de peau.

a b

Figure. 1.2: (a) et (b) représentent respectivement l'image originale et l'image segmentée après détection de la peau.

De nombreuses applications de reconnaissance de formes s'appuient sur les techniques de détection de la peau. Celles-ci constituent une étape préliminaire nécessaire pour des problèmes tels que la détection de visages, des mains ou du corps humain d'une façon générale. Cela a permis de réaliser des progrès notables dans plusieurs domaines qu'on abordera dans la section suivante.

Il existe trois approches majeures pour détecter les régions de peau dans une image: la détection basée sur la couleur [199] [221], la détection basée sur l'extraction des traits caractéristiques [28] [162] et la détection basée sur le mouvement [137]. Cette dernière est souvent utilisée conjointement avec l'information de couleur de peau [180] ou couplée avec un système basé sur les traits caractéristique de la peau [224] [1].

Nous présenterons d'une manière générale les approches basés sur l'extraction des traits caractéristiques de la peau et sur le mouvement dans les sections qui suivent avant de nous consacrer entièrement aux approches basées sur la couleur qui ont un lien direct avec notre travail.

1.1.2 Pourquoi détecter la peau

La détection de la peau est un processus qui permet de discriminer les pixels peau de ceux de non peau dans des images numérisées ou vidéos données. Il existe des capteurs (caméras, appareils de détection) qui opèrent dans le spectre électromagnétique non visible comme les systèmes de détection thermique [110]. Ces derniers se basent sur la détection des rayonnements infrarouges émis par le corps humain et sont couramment utilisés par les corps de défense et de protection civile dans les pays industrialisés [110]. Ne faisant pas parti de notre travail, nous omettons de détailler ce type de systèmes pour nos focaliser sur la détection de la peau dans des images couleur acquises par des appareils qui opèrent dans le spectre électromagnétique visible possédant des ondes électromagnétiques dont les longueurs d'onde sont comprises entre environ 380 et 780 nanomètres (nm) [207].

La détection de la peau humaine dans des images couleurs est considérée comme l'une des tâches les plus importantes pour relever le défi en surveillance et en reconnaissance de formes. La faculté que possède l'être humain en ce qui concerne la précision pour détecter et même identifier les personnes même dans des conditions défavorables est des plus étonnantes. Un changement dans les conditions d'éclairage même intense n'affectera pas le processus de reconnaissance par le système visuel humain [137].

Posséder des systèmes automatisés pouvant égaler la capacité visuelle de l'être humain résoudrait beaucoup de problèmes dans divers domaines. Beaucoup de chercheurs ont relevé le défi et se sont penchés sur cette question afin de faire de la modélisation de la peau un domaine de recherche à part. Se consacrer à ce domaine de recherche est motivé principalement par la multiplicité et l'émergence de nombreux champs d'applications tels que la vidéo surveillance [123] [130] [50], l'identification et l'authentification de personnes [128] [228], ainsi que les interfaces intelligentes homme-machine [203] [158].

On observe durant cette dernière décennie un besoin croissant pour des systèmes automatiques d'identification de personnes. On cite à titre d'exemple les besoins relatifs aux contrôles des frontières ainsi qu'à la surveillance des lieux publics tels que les banques, les aéroports et les centres commerciaux [96]. A cet égard, le gouvernement canadien a investi dans la sécurité publique 8.7 milliards de dollars sur cinq ans depuis 2002 [29] [30]. Ainsi, on peut noter que l'aéroport Pearson de Toronto possède déjà son système de reconnaissance d'individus [110].

L'identification d'une personne peut être réalisée à partir d'une image de l'individu, plus particulièrement de son visage. En plus de la forme et de la couleur, la signature acquise par une camera des vaisseaux sanguins qui se trouvent en dessous de la peau et de la rugosité de la surface et sa couleur contribuent à distinguer la peau des restes des objets de l'univers [42], identifier une *race*, estimer l'âge de la personne et même sa santé [16]. On sous-entend par *race* un regroupement d'êtres humains, qui se distingueraient par des traits physiques communs, généralement la couleur de

leur peau. On dénombre par exemple la race caucasienne, la race jaune et la race noire [86].

Pour ce faire, la reconnaissance de personnes nécessite l'acquisition, le traitement et l'interprétation de ces images. Afin de reconnaitre une personne, il faut avant tout localiser les parties du corps visibles, les analyser afin d'extraire des caractéristiques importantes qui seront utilisées pour interroger une base de données établie préalablement dans le but d'identifier une personne [228] [128].

Outre l'identification, un système de détection et de reconnaissance de visages peut également être utilisé pour faciliter la recherche et la navigation dans une masse de vidéos [3]. Un système pareil doit, au préalable, s'appuyer sur un système de détection et de segmentation des régions de peau.

Une autre application est le filtrage des images sur le Web [77] [229] [96]. En effet, le Web apparaît comme un immense gisement d'informations dont le libre accès conduit à des usages indésirables tels que la pornographie adulte et juvénile. Il existe des outils de filtrage de sites pour des applications comme le contrôle parental [96]. Ces derniers s'appuient dans le processus d'identification non uniquement sur l'information textuelle mais aussi sur une l'analyse d'images.

Situer et détecter les parties du corps apparentes représente une étape cruciale dans tous ces champs d'application. Cependant, il s'agit d'une tâche difficile à réaliser à cause de certains facteurs comme la diversité de la couleur de la peau et des conditions d'illumination variantes que nous présenteront dans la section qui va suivre.

1.2 Problèmes sous jacents à la détection de la peau

La plupart des efforts de recherches sur la modélisation de la peau se sont concentrées sur des systèmes de détection qui opèrent dans le spectre électromagnétique visible correspondant au domaine de la vision humaine. Bien que les êtres humains possèdent un système visuel développé qui possède la faculté de détecter et même d'identifier les parties correspondant au corps humain dans une image ou une scène sans beaucoup de peine, construire un système automatique qui accomplit de telles tâches représente un sérieux défi. Ce défi est d'autant plus grand lorsque les conditions d'acquisition des images sont très variables [2]. Toute la difficulté particulière de ce type de détection automatique émane principalement des facteurs suivants:

1.2.1 Changement d'illumination

L'apparence de la peau dans une image varie énormément en fonction de l'illumination de la scène lors de la prise de vue (voir figure 1.3). Les variations d'éclairage rendent la tâche de détection de la peau très difficile et peut entraîner une mauvaise classification des pixels de l'image d'entrée. Un changement d'intensité de la source lumineuse et donc dans le niveau d'illumination entraine un changement de la distribution colorimétrique de la peau dans les images couleur. La diversité des

conditions d'illumination est causée soit par l'environnement d'acquisition des images (images d'intérieur, d'extérieur), la nature de la source lumineuse (haute illumination, illumination colorée) ou bien les effets de l'ombrage dans les images. La variation de l'illumination est le problème le plus important traité actuellement qui dégrade sérieusement les performances de détection. Ce type de problème est connu sous le nom du *problème de constance de couleur* [119]. La détection de la peau dans un environnement non contrôlé reste donc un domaine de recherche ouvert.

Figure. 1.3: Différents éclairages.

1.2.2 Diversité ethnique

La couleur de la peau varie d'une personne à une autre appartenant à divers groupes ethniques et issues de différentes régions. Par exemple, la couleur de la peau de personnes appartenant à des races ethniques asiatiques, africaines ou Caucasiennes diffère considérablement d'un groupe à l'autre et varie du blanc, jaune au noir (figure 1.4) [119].

Figure. 1.4: Diversité ethnique.

1.2.3 La complexité de l'arrière plan

La richesse et la complexité de certaines images en ce qui concerne l'arrière plan telles que les images extérieurs entraine une détection plus difficile des pixels appartenant à l'image vu la diversité des couleurs existantes dans l'environnement (voir figure 1.5). Certaines de ces dernières possèdent même une distribution colorimétriques similaire ou presque à celle de la peau, d'où toute la difficulté de

détection correcte. Plus le décor de l'arrière plan est complexe plus la probabilité de tomber sur ces couleurs s'accroit [96].

Figure. 1.5: Décors complexe.

1.2.4 Présence ou absence des composants structurels

La présence des composants structurels tels que la barbe, la moustache, ou bien les lunettes peut modifier énormément les caractéristiques faciales telles que la forme, la couleur, ou la taille du visage causant ainsi une défaillance du système de détection. Par exemple, le port de lunettes ne permet pas de bien distinguer la partie correspondant aux yeux et une moustache ou une barbe modifie la forme du visage [2].

1.2.5 Les caractéristiques individuelles

Les caractéristiques individuelles correspondent aux particularités propres des personnes qui les différencient des autres individus comme l'âge et l'état de santé qui agit principalement sur la couleur de la peau.

La couleur de la peau peut varier considérablement durant la durée de vie d'une personne. Par exemple, dans les pays européens, les enfants les plus jeunes possèdent généralement un teint rougeâtre dans les parties à forte circulation sanguine comme les joues et plutôt clair ailleurs. Avec l'âge, la peau a tendance à prendre des couleurs soit plus sombres soit plus jaunâtre. L'état de santé des personnes influe également sur la détection de la peau. Une pigmentation de peau causant un changement de la couleur de la partie du corps atteinte empêche cette partie de la peau d'être détectée. Certaines maladies également provoquent d'importants changements de couleur de peau. Une autre caractéristique individuelle à ne pas négliger est l'effet des tatouages sur la peau. En effet, cela procure à la peau différentes couleurs en fonction des préférences des personnes tatouées et celles-ci sont distinctes de la couleur de la peau. D'autres facteurs peuvent aussi dégrader la détection de la peau comme le maquillage et les mouvements dans l'image traitée.

Beaucoup de problèmes rencontrés dans le cadre de la détection de la peau dans les images couleurs sont résolus en ayant recours à l'utilisation du spectre

électromagnétique non visible comme les infrarouges. Les méthodes reposant sur ce type de détection sont invariantes et beaucoup plus stables faces à toutes les difficultés de détection que nous venons de citer. Cependant, le prix prohibitifs des équipements nécessaires pour ces méthodes et les procédures de configuration fastidieuses ont limité leur utilisation à des domaines d'applications spécifiques comme le biomédical. A cet effet, nous nous concentrons exclusivement à des techniques de détection basées sur le spectre électromagnétique visible applicables à des images couleur 2D.

La détection de la peau dans les images couleur est confrontée à d'autres problèmes comme l'identification du meilleur espace de couleur et le choix d'une stratégie d'étiquetage de pixels de l'image qui font partie de la peau. Ces problèmes seront discutés ultérieurement.

1.3 Approches basées sur le mouvement et approches basées sur la géométrie et l'extraction de traits caractéristiques

Les approches basées sur le mouvement et l'extraction de traits caractéristiques ne se limitent pas uniquement à la détection des individus ou des parties du corps d'une manière générale. L'application des approches basées sur l'extraction des traits caractéristiques permet de détecter des parties du corps humains comme le visage et les mains en se basant sur les caractéristiques géométriques de celles-ci. En outre, elle permet une identification de la personne détectée en effectuant préalablement un apprentissage sur les caractéristiques physiques de celle-ci donc nécessite un traitement plus approfondi. L'application directe des approches basées sur le mouvement permettent en plus de la détection le suivi de mouvement dans des vidéos.

1.3.1 Approches basées sur le mouvement

L'utilisation de l'information de mouvement peut être un moyen simple pour mettre en œuvre une technique rapide de détection de peau dans une vidéo. Cette catégorie de systèmes suppose généralement que l'arrière plan de la scène vidéo est stationnaire et que les régions contenant la peau tels que le visage ou/et les mains par exemple sont en mouvement. Dans ce cas, ces régions peuvent être détectées par une simple différence entre l'image courante et l'image précédente. Il semble évident que les hypothèses retenues sont trop fortes. C'est pourquoi l'information de mouvement n'est jamais utilisée seule pour la détection. On la trouve utilisée conjointement avec l'information de couleur de peau ou couplée à un système basé sur la reconnaissance de traits caractéristiques [96].

L'exploitation de l'information de mouvement oriente la détection de peau vers des zones préférentielles en éliminant les zones sans intérêt et permet donc une forte réduction de la complexité de la détection et par conséquent une limitation importante des calculs et des traitements [137]. Cependant, les performances de ces systèmes sont fortement réduites lorsque la scène vidéo contient de nombreux objets en mouvement.

Parmi les techniques envisageables, on trouve la soustraction de l'arrière-plan (SAP) et la technique basée sur la différence entre deux images consécutives.

1.3.1.1 Soustraction de l'arrière plan (SAP) par modélisation statique

La soustraction de l'arrière-plan (SAP) forme la plus grande famille de méthodes de détection du mouvement. Celles-ci sont par ailleurs assez répandues et ont été utilisées dans de nombreux systèmes ([71], [10], [60], [102], [111] et [204]). Ceci s'explique probablement par leur simplicité théorique et leur faible complexité algorithmique. Le principe fondamental repose sur une estimation statistique de la scène observée. Le mouvement est détecté en comparant une image test avec le modèle d'arrière-plan calculé auparavant. Certaines hypothèses de base doivent par contre être respectées pour un fonctionnement adéquat de cette méthode. Tout d'abord, la caméra utilisée est fixe et ne doit bouger à aucun moment. Une caméra à l'épaule est un bon exemple de situation où la SAP ne peut pas s'appliquer.

Le modèle statistique calculé lors de la phase d'initialisation est constamment mis à jour, lui permettant ainsi de s'adapter aux changements qui peuvent se produire dans la scène observée. Cette capacité d'adaptation est commune à toutes les techniques de SAP par modélisation statistique et leur confère un atout majeur. Par ailleurs, cette méthode connaît plusieurs implantations différentes qui varient principalement selon le type de capteur utilisé [96].

Visible (2D) est la première catégorie de méthodes de soustraction d'arrière-plan qui regroupe les techniques basées sur l'utilisation d'images 2D dans le spectre visible [96]. La technique de base consiste à modéliser l'arrière-plan à partir de plusieurs images acquises séquentiellement. Pour chacun des canaux (R, G et B), une moyenne et une variance sont calculées. Lorsqu'un pixel test doit être classifié, il faut tout d'abord lui soustraire la moyenne correspondante dans le modèle statistique. Il sera alors étiqueté comme un pixel contenant du mouvement seulement si la valeur absolue du résultat dépasse un certain multiple de l'écart type correspondant. Par ailleurs, d'autres modèles de couleur ont été utilisé auparavant pour la modélisation statistique, comme par exemple le YUV par Wren et al. [217] et le HSV par Cucchiara et al. [60].

Horprasert et al. [102] ont proposé un nouveau modèle de couleur basé sur l'espace RGB. Leur technique permet la classification des pixels en quatre catégories: appartenant à l'arrière plan original, illuminé, ombré, et en mouvement. En pratique, certaines erreurs de classification peuvent également se produire entraînant, par exemple, l'identification d'un objet en mouvement comme étant de l'ombre [183].

Il y a finalement un très grand nombre de méthodes de SAP par modélisation statistique qui ne sont pas abordées ici [10], notamment pour des raisons de complexité et pour lesquelles les gains en performance sur la technique de base sont relativement négligeables.

1.3.1.2 Différence entre deux images consécutives

Etant peu complexe, la différence entre deux images consécutives représente une solution très intéressante. Comme son nom l'indique, elle consiste à soustraire une image acquise au temps t_n d'une autre au temps t_n+k, où k est habituellement égal à 1. Ainsi, l'image résultante sera vide si aucun mouvement ne s'est produit pendant l'intervalle de temps observé car l'intensité et la couleur des pixels seront presque identiques. Par contre, si du mouvement a lieu dans le champ de vue, les pixels frontières des objets en déplacement devraient changer drastiquement de valeurs, révélant alors la présence d'activité dans la scène. Cette technique nécessite très peu de ressources, car aucun modèle n'est nécessaire. Cela implique donc qu'il n'y a pas de phase d'initialisation obligatoire avec une scène statique, ce qui procure une très grande flexibilité d'utilisation. De plus, une opération de soustraction d'images requiert très peu de puissance de calcul, lui conférant un avantage supplémentaire [173]. Cependant, les résultats obtenus avec cette méthode ne sont pas aussi intéressants que ceux générés en utilisant un modèle statistique de l'arrière plan.

Les techniques basées sur le mouvement, telle que la méthode de différence d'image sont des techniques simples permettant de faire rapidement une estimation de la position d'un objet en mouvement. Cette estimation permet de réduire la zone de recherche. On trouve cette approche souvent utilisée conjointement avec l'information de couleur de peau ou couplée avec un système basé sur la reconnaissance de traits caractéristiques. Cependant cette technique impose des contraintes sur l'environnement: 1) La caméra doit être fixe sous peine de détecter l'image entière comme objet en mouvement [216], 2) Les sources lumineuses doivent être constantes et fixes. Un changement de luminosité entraîne la détection de la zone de changement comme étant en mouvement. Celui-ci peut être provoqué par l'allumage d'une lampe, le passage d'un individu créant une ombre ou le passage d'un nuage [48].

1.3.2 Approches basées sur la géométrie et l'extraction de traits caractéristiques

Les méthodes qui se basent sur ce type d'approche sont en général appliquées pour la segmentation des régions de peau des visages ou des mains pour des fins de reconnaissance de personnes et de gestes. Cela permet de localiser une personne, de la différencier des autres intervenants ou d'interpréter l'un de ses gestes. Cette segmentation exige la définition de modèles pour les traits caractéristiques, voir même un modèle du corps humain plus ou moins détaillé.

Les traits caractéristiques recherchés dans le domaine de la détection de visage par exemple peuvent être les yeux, les contours extérieurs, le nez et la bouche que l'on associe à des configurations ("templates") connues à priori ou apprises [45]. Grâce à l'application de contraintes géométriques, il est désormais possible de déterminer le positionnement relatif de ces traits caractéristiques et donc d'obtenir une information sur la présence ou non d'un visage.

Govindaraju et al. [91] a proposé un système basé sur les contours. Il se base sur une modélisation du visage sous forme d'agencement de trois courbes: sommet, coté droit et coté gauche du visage. La détection et l'extraction de ces trois types de courbes et leur regroupement d'après leur position permet de connaitre la position du visage. Cependant, L'utilisation de contours actifs permet une adaptation aux variations de forme du visage et conduit ainsi à une meilleure délimitation du visage et constituent les méthodes basées sur le contour les plus abouties telles que celles proposées par Waite et al. [214], Craw et al. [58] et Cootes et al. [54]. Il existe une autre approche basée sur les contours qui consiste en une extraction des contours de l'image et leur mise en correspondance avec une ellipse modélisant un visage [112] [174]. Les traits du visage pris en compte le plus souvent sont les yeux, les sourcils, la bouche et le nez qui peuvent à leur tour être détaillés à un niveau encore plus fin. Cela suppose que les caractéristiques du visage ont des positions fixes. Il existe un nombre important de caractéristique définissables et de techniques de détection d'où la diversité des méthodes qui se basent sur l'exploitation des traits caractéristiques.

L'un des premiers travaux enregistrés dans ce cadre est celui de Kanade en 1973 [120] [125]. La méthode qu'il a élaborée a constitué un premier pas vers les techniques de détection et de reconnaissance de visage. Le système est capable de retrouver le positionnement des yeux, de la bouche, du nez, ainsi que les contours du visage. Dans le cas où le profil obtenu ne correspond pas au modèle de référence, l'image est considérée comme n'étant pas un visage. La méthode la plus utilisée pour extraire les caractéristiques du visage est la définition d'un masque de caractéristiques générique et d'établir entre celui-ci et l'image à analyser une certaine corrélation (Sumi et Otha [192], Zelinsky et Heinzmann [227]).

D'autres méthodes plus généralisées ont été proposées pour l'extraction des paramètres du visage (Leung et al. [138], Graf et al. [93], Burl et al. [31]). Celles-ci sont peu robuste et sont sujet à plusieurs fausses alertes. Afin d'améliorer la robustesse du système, les positions relatives des traits du visage détectés sont prises en compte par des techniques de comparaison avec un modèle de déformation [174]. La déformation du visage est due aux expressions faciales. Etant donné que l'expression faciale modifie l'aspect du visage, elle entraîne forcément une diminution du taux de reconnaissance. L'identification de visage avec expression faciale est un problème difficile qui est toujours d'actualité et qui reste non résolu [2].

Figure. 1.6: Exemple de variation d'expressions.

Les approches basées sur l'extraction de traits du visage sont plus efficaces que les approches basées sur les contours car elles analysent des structures locales de l'image conjointement à des modèles géométriques de visages [174].

Le même type d'approche est utilisé pour la modélisation et le suivi des mains. Tandis que certains auteurs proposent une forme 2D de la main, d'autres penchent plutôt pour une modélisation générique 3D, où la main est représentée par un modèle de type volumique articulé (figure 1.7).

Figure. 1.7: Modèle 3D de la main.

L'animation de ce modèle peut être effectuée selon 6 degrés de liberté pour le positionnement global et 5 degrés de liberté pour chaque doigt et cela conformément au paramétrage MPEG-4 [168]. Il est nécessaire d'ajuster préalablement les paramètres morphologiques du modèle afin d'assurer la robustesse de la procédure de suivi de la main dans des séquences. Cette opération est effectuée à partir d'une seule image de la main ouverte, les doigts écartés. On calcule les paramètres de la morphologie des doigts (rayon et longueur) et de la paume (longueur et largeur) à partir des dimensions des régions correspondantes [141].

Les objectifs que nous avons visés nous ont amené à choisir une méthode basée sur la détection de peau par une approche couleur. En effet, nous avons voulu que notre modèle de peau soit adapté pour les images couleurs fixes. De ce fait, Les approches basées sur le mouvement et l'extraction de traits caractéristiques ne correspondent pas à notre besoin. A la différence de beaucoup de travaux qui ne traitent que des parties spécifiques d'un corps humain telles que les visages ou que les mains, nous visons ici un modèle général et robuste pour la détection de régions de peau dans une image par rapport à la diversité d'applications, de conditions de lumière et la diversité ethnique. Ainsi, il nous est apparu qu'une méthode basée sur la détection de peau par une approche couleur est la plus adaptée à condition de bien choisir un espace de couleur approprié à la détection de peau. Nous présenterons dans la section suivante les différents espaces de couleur adaptés à la détection de la peau.

1.4 Apport des espaces de couleurs pour la détection de la peau

La couleur est une propriété importante associée aux objets. Elle fournit une information utile pour ces objets afin de faciliter leur localisation, identification et leur interprétation. On identifie aisément son équipe favorite lors d'un match de football en se basant sur la couleur de la tenue et on reconnait facilement un fruit mur apte à la consommation en examinant sa couleur.

La couleur est un phénomène psycho-physiologique provoqué par l'excitation de photorécepteurs situés sur la rétine de l'œil par une onde électromagnétique. Elle est le résultat conjugué de [57]:
-la source lumineuse qui éclaire la scène;
-la géométrie d'observation (angles d'éclairement et d'observation);
-la scène et ses caractéristiques physiques;
-l'œil de l'observateur ou le capteur de la caméra;
-le cerveau de l'observateur.

1.4.1 La théorie trichromatique

L'origine de la théorie trichromatique revient à 1672 avec l'expérience de Newton sur la dispersion de la lumière à travers un prisme. Par la suite Thomas Young a suggéré en 1801 que trois couleurs primaires étaient suffisantes pour produire toutes les couleurs de façon additive. Helmotz a poursuivis ces travaux et a pu prouver la théorie trichromatique en 1960 en découvrant trois types de récepteurs dans la rétine correspondant aux trois types de teinte L, M et S. Leur réponse maximale se situe respectivement dans les teintes bleues à 440 nm pour les cônes de type S (*Short*), dans les teintes vertes à 545 nm pour les cônes de types M (*Medium*) et dans les teintes rouges à 580 nm pour les cônes de type L (*Long*).
Il serait alors intéressant de représenter les couleurs dans un espace tridimensionnel dont les couleurs de base correspondent aux couleurs primaires. Ainsi, un vecteur couleur [*S*] est représenté par une combinaison linéaire des vecteurs de la base ([*R*], [*G*], [*B*]):

$$[S] = r[R] + g[G] + b[B] \qquad (1.1)$$

Où: r, g et b correspondent aux coordonnées des composantes trichromatiques et représentent les quantités respectives des primaires utilisées. Chaque pixel d'une image couleur contient donc les trois composantes qui définissent sa couleur (figure 1.9). C'est le principe de la vision trichromatique [96].

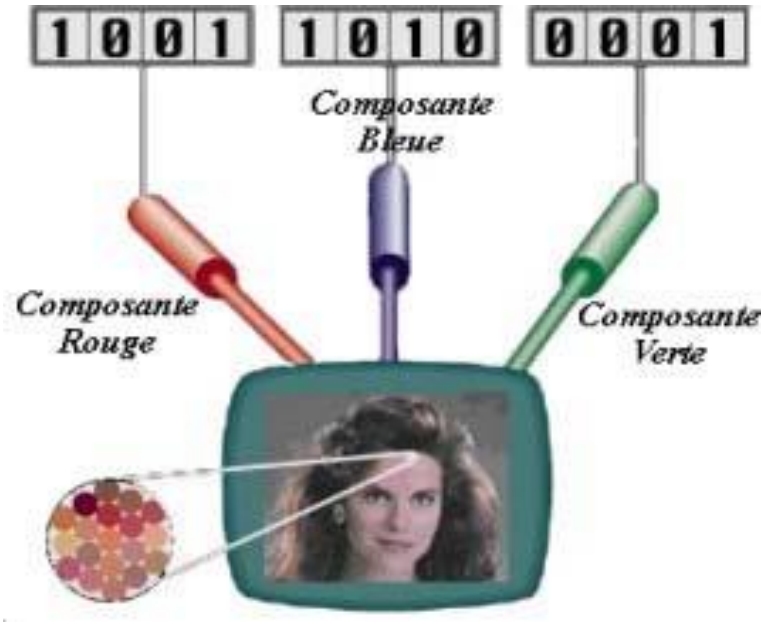


Figure. 1.8: Représentation de la couleur.

Tous les travaux effectués dans le but d'obtenir les fonctions colorimétriques permettant de calculer les composantes trichromatiques d'images soumises à diverses conditions d'illumination ont donné naissance au standard défini par la CIE (Commission Internationale de l'Eclairage) en 1931 [41]. Nous allons présenter dans la section suivante quelques espaces de représentation de la couleur.

La couleur constitue un moyen simple et efficace pour la distinction entre les pixels de peau et ceux de non peau. Il a été montré que la couleur de la peau est située dans un intervalle étroit de l'espace de couleur [77] [78] et [124]. Il est donc possible de distinguer les pixels peau de ceux de non peau. De nombreux chercheurs se sont intéressés sur cet aspect et se sont occupés à améliorer le processus de détection de peau en exploitant différents espaces de couleur.

1.4.2 Les espaces de représentation de la couleur

Il existe de nombreux systèmes de représentation de la couleur: les systèmes primaires, les systèmes luminance-chrominance, les systèmes perceptuels, les systèmes d'axes indépendants, etc. Bien qu'il existe une forte dépendance au système de primaires RGB en raison notamment de la dépendance aux matériels (cartes d'acquisition, cartes vidéo, caméras, écran, ...) qui sont conçus pour faire des

échanges d'informations en se basant uniquement sur le triplet (R,G,B), les systèmes les plus utilisés et les plus répondus sont les systèmes de type luminance-chrominance et les systèmes perceptuels.

1.4.2.1 Espaces de primaires

Il existe une multitude d'espaces (R,G,B) qui dépendent des primaires utilisées pour reproduire la couleur. Nous présentons ici l'espace (R_C,G_C,B_C) de la CIE.

1.4.2.1.1 L'espace (R_C,G_C,B_C) de la CIE

Cet espace a été défini en 1931. L'indice C correspond à une référence pour la CIE (Commission Internationale de l'Eclairage). Chaque stimulus de couleur est représenté par un point C qui définit le vecteur couleur $\vec{OC}$. Les coordonnées de ce vecteur sont les composantes trichromatiques R_C, G_C et B_C. Les points correspondant à des stimuli de couleur, dont les composantes trichromatiques sont positives, sont contenus dans un cube, connu sous le nom de cube des couleurs (figure 1.9). L'origine O correspond au noir ($R_C = G_C = B_C = 0$) tandis que le blanc de référence est défini par le mélange unitaire des trois primaires ($R_C = G_C = B_C = 1$). La droite passant par les points Noir et Blanc est appelée axe des gris, axe des couleurs neutres ou encore axe achromatique. En effet, les points de cette droite représentent des nuances de gris allant du noir au blanc [207].

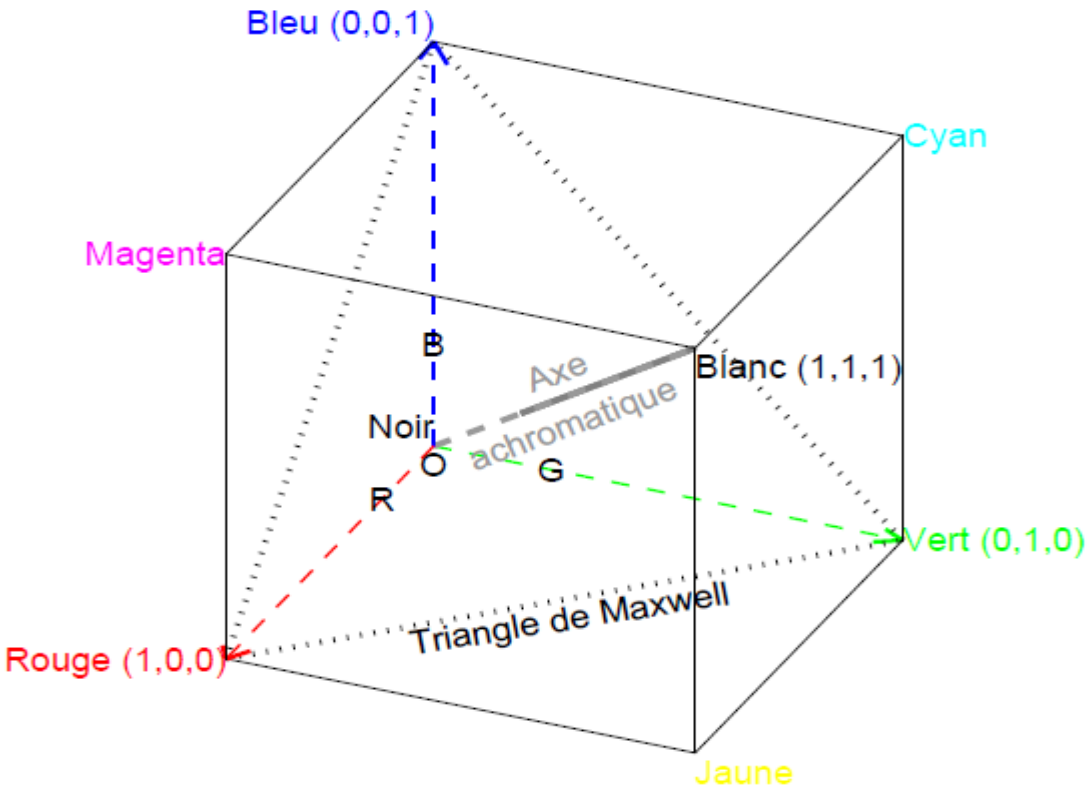


Figure. 1.9: Cube des couleurs [207].

1.4.2.1.2 L'espace (r_C,g_C,b_C)

Les composantes trichromatiques d'un stimulus de couleur du système (R_C,G_C,B_C) sont liées à sa *luminance* (nous définissons ici la luminance comme étant l'attribut d'une sensation visuelle selon laquelle une surface paraît émettre plus ou moins de lumière).

Deux stimuli de couleur peuvent ainsi posséder le même caractère chromatique, que nous appellerons chrominance, mais avoir des composantes trichromatiques R_C, G_C et B_C différentes et ceci à cause de leur luminance. Afin d'obtenir des composantes qui ne tiennent compte que de la chrominance, il convient donc de normaliser les valeurs des composantes trichromatiques par rapport à la luminance. Ceci est réalisé en divisant chaque composante par la somme des trois. Les composantes ainsi obtenues sont appelées coordonnées trichromatiques, coordonnées réduites ou encore composantes normalisées. Pour l'espace (R_C,G_C,B_C) de la CIE, elles sont notées r_C,g_C,b_C et sont définies par:

$$\begin{cases} r_C = \dfrac{R_C}{R_C + G_C + B_C} \\ g_C = \dfrac{G_C}{R_C + G_C + B_C} \\ b_C = \dfrac{B_C}{R_C + G_C + B_C} \end{cases} \qquad (1.2)$$

L'espace de représentation associé aux coordonnées trichromatiques est appelé l'*espace* (R_C,G_C,B_C) normalisé. Il est noté (r_C,g_C,b_C). Comme $r_C + g_C + b_C = 1$, deux composantes suffisent à représenter la chrominance d'une couleur. De cette manière, Wright et Guild ont proposé un diagramme appelé diagramme de chromaticité (r,g) représenté sur la figure 1.10 [207]. On note sur cette figure l'existence de coordonnées négatives. Ces coordonnées représentent des couleurs physiquement non réalisables par synthèse additive. En réponse à ce problème, entre autres, la CIE a établi le système de référence colorimétrique (X,Y,Z).

1.4.2.1.3 L'espace (X,Y,Z) de la CIE

Les primaires *X*, *Y* et *Z*, dites primaires de références, ont été créées afin de pouvoir exprimer toutes les couleurs par des composantes trichromatiques positives. La primaire Y est une information sur la luminosité (plus précisément Y représente la luminance visuelle). On peut noter qu'il est possible de passer de n'importe quel espace (*R*,*G*,*B*) à l'espace (*X*,*Y*,*Z*) par l'intermédiaire d'une matrice de passage dont les coefficients sont principalement conditionnés par le choix des primaires *R*, *G* et *B* utilisées. Pour l'espace (R_C,G_C,B_C), la CIE a défini les coordonnées de l'espace (X,Y,Z):

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.618 & 0.177 & 0.205 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.056 & 0.944 \end{pmatrix} * \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (1.3)$$

1.4.2.1.3 L'espace xyz

De même que pour l'espace (*R*,*G*,*B*), on définit une normalisation des valeurs des composantes trichromatiques pour l'espace (X,Y,Z) comme suit:

$$\begin{cases} x = \dfrac{X}{X+Y+Z} \\ y = \dfrac{Y}{X+Y+Z} \\ z = \dfrac{Z}{X+Y+Z} \end{cases} \quad (1.4)$$

Comme *x*+*y*+*z* = 1, *z* peut être déduit à partir de *x* et *y*. cela permet de représenter la couleur dans un plan et donc de construire le diagramme de chromaticité (*x*, *y*), illustré par la figure 1.11. Il est notable que sur ce diagramme que toutes les couleurs sont exprimées par des coordonnées trichromatiques positives, contrairement au diagramme de la figure 1.10 [207].

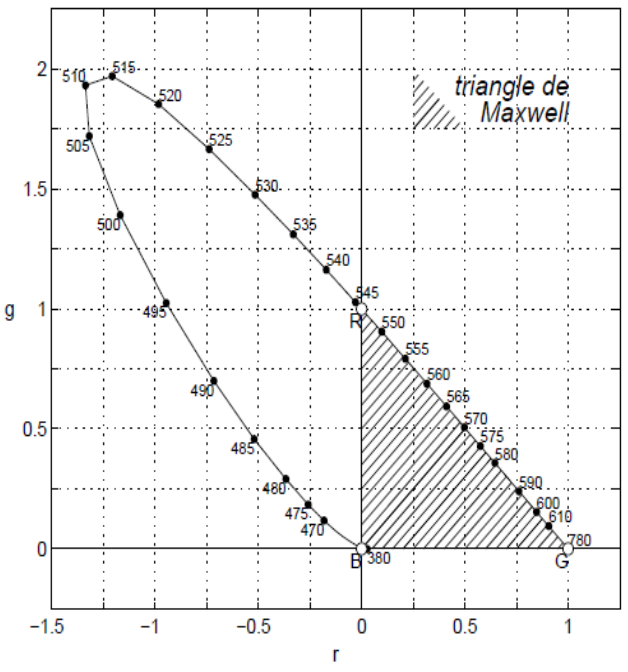


Figure. 1.10: Diagramme de chromaticité (r_C,g_C) de la CIE [207].

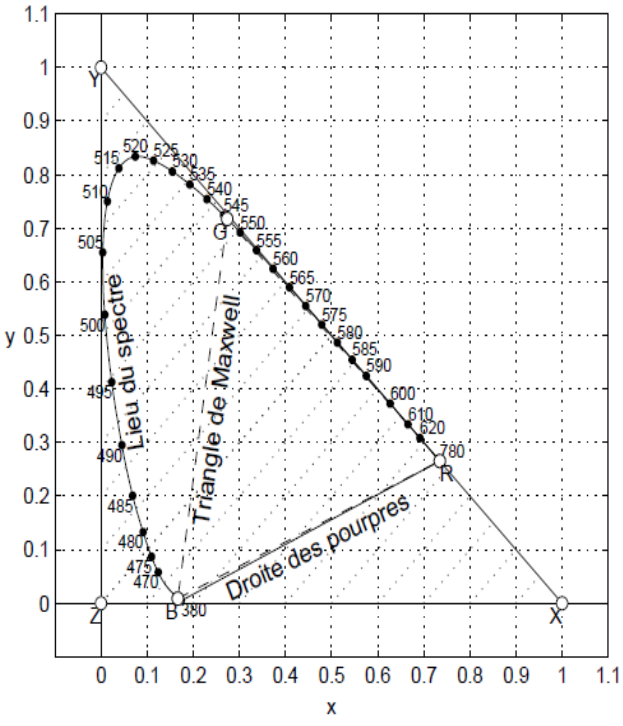


Figure. 1.11: Diagramme de chromaticité (*x*, *y*) [207].

1.4.2.2 Les systèmes luminance-chrominance

Les espaces luminance-chrominance possèdent une composante appelée *luminance* qui permet de quantifier la luminosité, et deux autres composantes de chrominance notées (*Chr*1 et *Chr*2) qui sont suffisantes pour quantifier le caractère chromatique d'un stimulus de couleur. Les composantes d'un système luminance-chrominance sont évaluées à partir des composantes trichromatiques *R*, *G* et *B*, soit par une transformation linéaire définie par une matrice de passage, soit par une transformation non linéaire. Le type de transformation est lié à la nature même du système. Il existe différents espaces luminance-chrominance. On trouve, entre autres, les espaces perceptuellement uniformes (où la luminance est notée *L*), les espaces antagonistes (où la luminance est notée *A*) ou encore les espaces de télévision (où la luminance est notée *Y*).

1.4.2.2.1 Les espaces de télévision

Le codage des signaux de télévision en couleur a été réalisé de façon à rester compatible avec les téléviseurs noir et blanc qui doivent pouvoir recevoir en noir et blanc les émissions en couleur. De même, les téléviseurs couleurs doivent pouvoir recevoir en noir et blanc les émissions en noir et blanc. Pour satisfaire ces deux principes, les signaux de télévision séparent donc l'information de luminance de celle de chrominance. Cette séparation peut être réalisée par une transformation linéaire des composantes trichromatiques du système (R_C,G_C,B_C). La luminance correspond à la composante *Y* du système (*X*,*Y*,*Z*). Les composantes de chrominance *Chr*1 et *Chr*2 sont alors calculées par les relations suivantes:

$$\begin{cases} Ch1 = a_1(R-Y) + b_1(B-Y) \\ Ch2 = a_2(R-Y) + b_2(B-Y) \end{cases} \tag{1.5}$$

Les coefficients a_1, b_1, a_2 et b_2 sont spécifiques aux différents standards de transmission (NTSC, PAL ou SECAM) et *Y* représente la luminance. Les principaux espaces de télévision sont le système YIQ qui correspond à la norme NTSC, le système YUV qui correspond à la norme PAL et le système YCbCr dédié au codage digital des images de la télévision numérique, correspond quand à lui à la norme ITU.BT-601 et fait partie du nouveau standard de compression JPEG2000 [196]. Les principales transformations sont données par les équations suivantes [184]:

$$\begin{bmatrix} \mathrm{Y} \\ \mathrm{Cb} \\ \mathrm{Cr} \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \begin{bmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{bmatrix} \tag{1.6}$$

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.289 & 0.436 \\ 0.212 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1.7)$$

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1.8)$$

1.4.2.2.2 Les systèmes perceptuellement uniformes

L'espace *(X,Y,Z)* n'est pas perceptuellement uniforme car il existe des zones dans le diagramme de chromaticité (*x*, *y*) où les différences de couleurs ne sont pas perceptibles par un observateur. Dans l'espace RGB également, deux couleurs perceptuellement semblables peuvent être loin l'une de l'autre. En effet, des couleurs perceptuellement proches peuvent correspondre à des écarts de couleurs importants dans l'espace de représentation adopté tandis que des couleurs perceptuellement très différentes peuvent correspondre à des écarts de couleurs trop faibles, d'où la nécessité dans certains cas d'utiliser des espaces de représentation perceptuellement uniformes. Les espaces *L*u*v** et *L*a*b** sont deux espaces perceptuellement uniformes définis par le CIE en 1976 [53]. L'information de luminance est dans ce cas appelée *clarté*. Le blanc de référence utilisé est alors caractérisé par ses composantes trichromatiques qui sont notées X^W, Y^W et Z^W respectivement pour les primaires *X*, *Y* et *Z*. La luminance est représentée comme suit:

$$L^* = \begin{cases} 116 \times \sqrt[3]{\frac{Y}{Y^W}} - 16 & si \ \frac{Y}{Y^W} > 0.008856 \\ 903.3 \times \frac{Y}{Y^W} & si \ \frac{Y}{Y^W} \le 0.008856 \end{cases} \quad (1.8)$$

Les composantes u* et v* de l'espace L*u*v* sont calculées comme suit:

$$u^* = 13 \times L^* \times (u' - u'^W) \quad (1.9)$$

$$v^* = 13 \times L^* \times (v - v'^W) \quad (1.10)$$

$$u' = \frac{4X}{X + 15Y + 3Z} \quad (1.11)$$

$$v' = \frac{9Y}{X + 15Y + 3Z} \quad (1.12)$$

avec u'^{W} et v'^{W} sont les composantes de chrominance de u' et v' correspondant au blanc de référence.
Les composantes de chrominance pour le système *L*a*b** sont:

$$a^* = 500 \times (f(\frac{X}{X^W}) - f(\frac{Y}{Y^W})) \quad (1.13)$$

$$b^* = 200 \times (f(\frac{Y}{Y^W}) - f(\frac{Z}{Z^W})) \quad (1.14)$$

avec :

$$f(x) = \begin{cases} \sqrt[3]{x} & si \quad x > 0.008856 \\ 7.787x + \frac{16}{116} & si \quad x \le 0.008856 \end{cases} \quad (1.15)$$

La figure 1.13 suivante illustre l'espace de couleur L*a*b*.

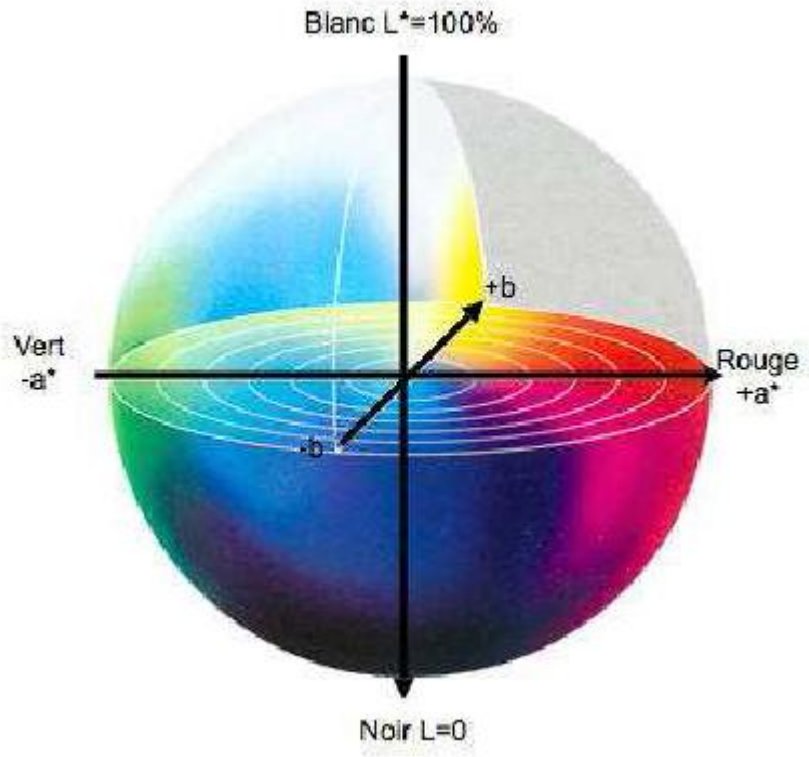


Figure. 1.12: L'espace de couleur CIELab [89].

La composante *L* représente la luminosité; sa plage de valeur varie entre 0 % pour la couleur noire et 100 % pour la couleur blanche. Les deux autres composantes *a* et *b* décrivent la couleur. La composante *a* permet de parcourir l'axe de couleur rouge-vert, pour une gamme de couleurs allant du vert (-128) au rouge (+127), et la

composante *b* parcourt l'axe de couleur jaune-bleu, pour une gamme de couleurs allant du bleu (-128) au jaune (+127) [184].

1.4.2.2.3 Les espaces antagonistes

Cette famille d'espaces de représentation de la couleur a été créée pour tenter de modéliser le système visuel humain. Afin de parvenir à ce but, quelques auteurs ont proposés différents modèles appliqués à l'analyse d'image. On retrouve notamment les travaux de Faugeras [73] (espace (*L*, *M*, *S*)) ou encore ceux de Garbay [81] (espace (*A*, *C*1, *C*2)). Ces travaux sont basés sur la théorie des couleurs opposées de Hering. Selon cette théorie, l'information couleur captée par l'œil est transmise au cerveau sous la forme de trois composantes, une composante achromatique (*L* ou *A*) et deux composantes de chrominance, correspondant respectivement à un signal d'opposition vert-rouge et à un signal d'opposition jaune-bleu. En ce sens les espaces (*L**,*u**,*v**) et (*L**,*a**,*b**) présentés précédemment peuvent être considérés comme des espaces antagonistes [207].

1.4.2.3 Les espaces perceptuels

Dans ce type d'espace, la couleur est décrite comme l'homme la qualifie, c'est-à-dire par rapport à la *luminosité*, la *teinte* et la *saturation*. La teinte correspond aux dénominations des couleurs telles que le rouge, vert, bleu, jaune, etc. La saturation, elle, est une grandeur permettant d'estimer le niveau de coloration d'une teinte indépendamment de sa luminosité. Les espaces perceptuels sont en fait des cas particuliers des espaces luminance-chrominance. Il existe de nombreux espaces de ce type dans la littérature, présentés sous différentes dénominations telles que (*I*,*S*,*H*), (*H*,*S*,*L*), (*H*,*S*,*I*), (*H*,*S*,*V*), (*L*,*T*,*S*), etc. On retrouve notamment les espaces de coordonnées polaires et les espaces de coordonnées perceptuelles [207].

1.4.2.3.1 Les espaces de coordonnées polaires

Cette famille se déduit directement des espaces luminance-chrominance dans lesquels, la représentation de la couleur se fait avec un axe pour la luminosité et un plan pour la chrominance comme le montre la figure 1.13. Un point *P* aura alors pour coordonnées *L*, *C* et *H*; *L* étant la luminosité, *C* le *chroma* (qui est une grandeur permettant d'estimer le niveau de coloration d'une teinte, tout comme la saturation, mais qui, lui, est dépendant de la luminosité) et *H* l'information de teinte. Les espaces de coordonnées polaires sont en fait une transposition des coordonnées cartésiennes des espaces luminance-chrominance en coordonnées polaires. *C* représente alors le module des coordonnées du point *P* et *H* l'angle d'orientation [207].

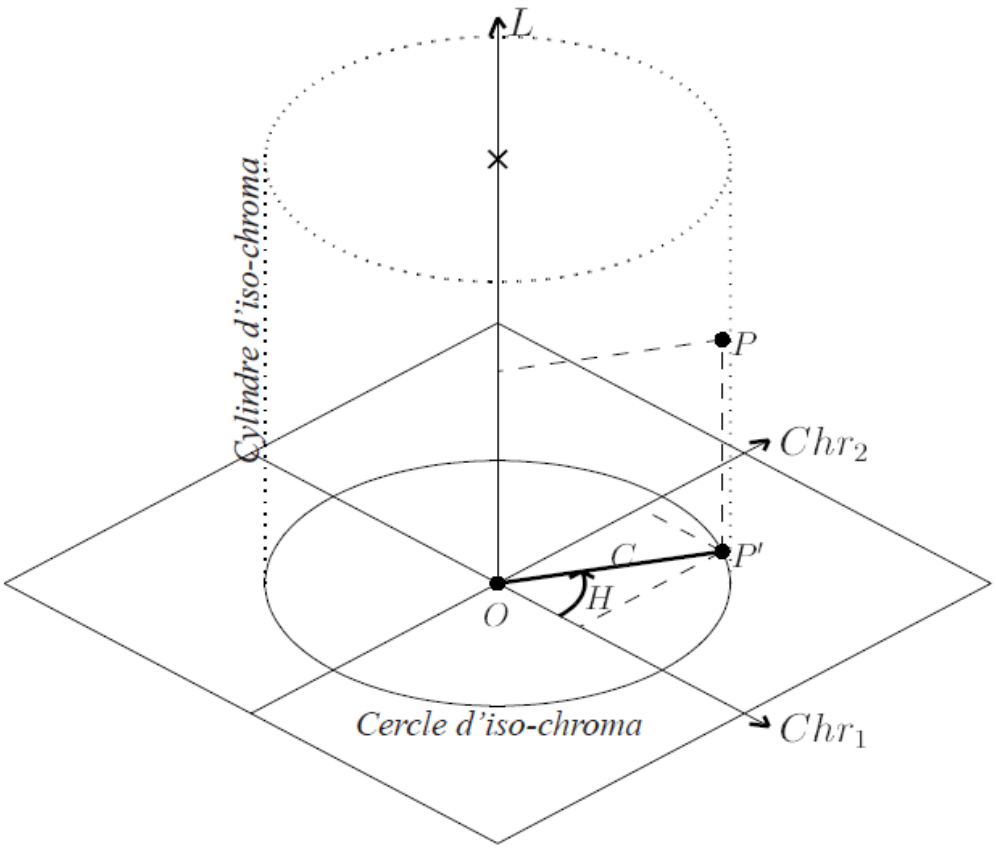


Figure. 1.13: Système de coordonnées polaires [207].

Il est possible de construire un système (*L*,*C*,*H*) à partir de n'importe quel système luminance-chrominance en utilisant les équations suivantes:

$$C = \left\| \overrightarrow{OP'} \right\| = \sqrt{chr_1^2 + chr_2^2} \tag{1.16}$$

$$H = \arctan\left(\frac{chr2}{chr1}\right) \tag{1.17}$$

En utilisant les équations 1.16 et 1.17, la CIE définit les composantes de systèmes de coordonnées polaires (*L*,*C*,*H*) à partir des systèmes uniformes (*L**,*u**,*v**) et (*L**,*a**,*b**) obtenant ainsi respectivement les systèmes (L^*_{uv},C^*_{uv},h_{uv}) et (L^*_{ab},C^*_{ab},h_{ab}) [53].

1.4.2.3.2 Les systèmes humains de perception de la couleur

Ces systèmes sont directement évalués à partir d'un système de primaires et représentent la couleur en termes d'intensité (*I*), de saturation (*S*) et de teinte (*T*). L'intensité correspond à l'information de luminosité mais elle est désignée ainsi principalement dans un souci de la différencier de celle des systèmes de coordonnées polaires vue précédemment. La saturation représente le niveau de

coloration d'une surface indépendamment de sa luminance, au contraire du chroma. Chroma et saturation sont ainsi liés par la relation: *S = C / L*. La teinte est notée *T* pour la différencier de la teinte notée *H* vue au paragraphe précédent. Ces trois composantes forment un système (*I,S,T*) et rencontrent des formulations très diverses dans la littérature. Parmi les plus connus de ces espaces sont l'espace HSI et HSV [207].

L'espace HSI est un système de représentation est issu de la rotation du cube des couleurs RGB. Cela consiste à faire pivoter le cube sur le coin représentant le noir; ainsi, l'axe achromatique constitue l'axe des intensités I et la couleur est définie par une position sur un pallier circulaire où la saturation S représente le rayon et la teinte H représente l'angle. La transformation de l'espace RGB à l'espace HSI sont données par [184]:

$$\begin{cases} I = \dfrac{R+G+B}{3} \\ S = 1 - \dfrac{3 * \min(R,G,B)}{R+G+B} \\ H = \arccos\left(\dfrac{0.5*(R-G)+(R-B)}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right) \end{cases} \tag{1.18}$$

L'espace HSV est un espace dérivé de l'espace RGB, le plus souvent, utilisé dans des applications informatiques de graphisme. Il a été formellement décrit en 1978 par Alvy Ray Smith [187]. Les couleurs dans cet espace sont représentées selon des notions de teinte (*Hue*), de pureté (*Saturation*) et de luminosité (*Value*).
La teinte caractérise la couleur. Sa valeur varie entre 0 et 360° et elle correspond à la position de la couleur dans un cercle où se positionne six couleurs primitives: rouge, jaune, vert, cyan, bleu et magenta et les différentes couleurs intermédiaires. La saturation correspond à la pureté de la couleur et indique la quantité de gris dans la couleur. Sa valeur varie entre 0 et 100 %. La luminosité correspond à la brillance perçue de la couleur. Elle varie entre 0 et 100 %. La figure suivante illustre l'espace de couleur HSV [207].

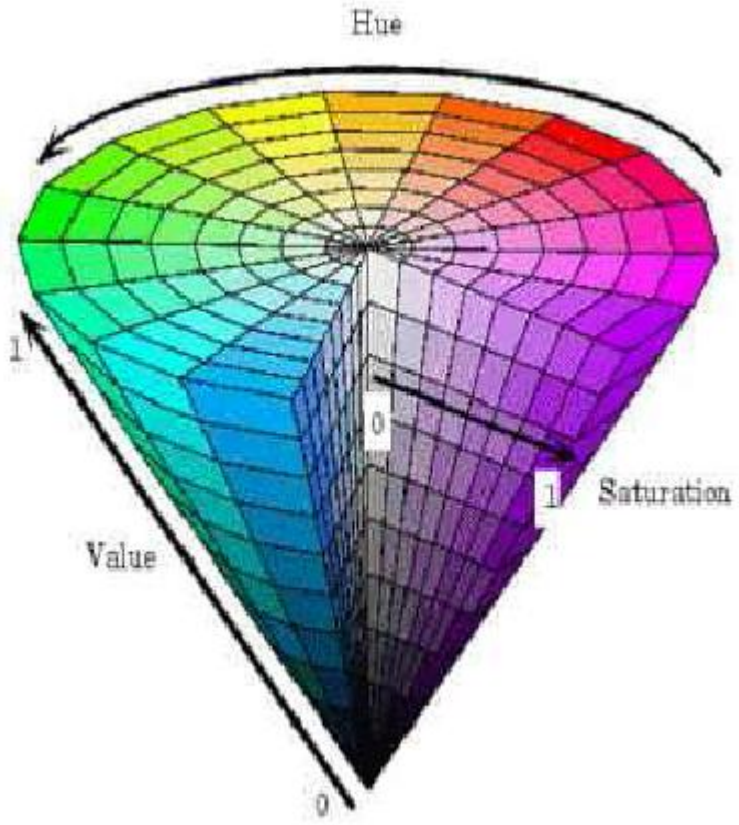


Figure. 1.14: Espace de couleur HSV [89].

Les valeurs des composantes HSV sont obtenues à partir d'une transformation non linéaire des couleurs RGB, selon les relations mathématiques suivantes [89]:

$$V = \max(R, G, B) \tag{1.19}$$

$$S = \frac{V - \min(R, G, B)}{V} \tag{1.20}$$

$$H = \begin{cases} \dfrac{G - B}{V - \min(R, G, B)} & si \quad V = R \\ 2 + \dfrac{B - R}{V - \min(R, G, B)} & si \quad V = G \\ 4 + \dfrac{\mathrm{R} - \mathrm{G}}{\mathrm{V} - \min(\mathrm{R}, \mathrm{G}, \mathrm{B})} & si \quad V = B \end{cases} \tag{1.21}$$

1.4.2.4 Les systèmes d'axes indépendants

Les trois composantes du système RGB sont fortement corrélées. C'est-à-dire qu'elles portent en elles une information commune. En effet, elles possèdent un fort facteur de luminance réparti sur chacune d'elles. Traiter chaque composante à part peut ainsi conduire à une perte d'information. Afin de palier à ce problème, de

nombreux auteurs se sont intéressés à déterminer des systèmes de représentation de la couleur dont les composantes sont indépendantes (composantes qui portent des informations différentes) en se basant sur l'analyse en composantes principales comme l'espace $I_1I_2I_3$ [184]. Il s'agit d'un système introduit par Otha et al. [157]. Son principe est de déterminer les trois axes de plus grande variance de l'ensemble des couleurs. Les auteurs parviennent à obtenir un système d'axes à partir d'un échantillon de huit images et ont déterminé une transformation linéaire menant de l'espace RGB à l'espace $I_1I_2I_3$ comme suit:

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & -1/2 \\ -1/4 & 1/2 & -1/4 \end{bmatrix} \begin{bmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{bmatrix} \tag{1.22}$$

L'espace $I_1I_2I_3$ fait également parti de la famille des systèmes de type de luminance-chrominance, du moment que I_1 correspond à la luminance et I_2 et I_3 aux composantes de chrominance.

1.4.3 Choix d'un espace de couleur pour la modélisation de la peau

Il existe une multitude d'espaces de couleurs. La liste exhaustive de ces espaces dépasse largement ceux cités dans la section précédente. Cependant, un choix approprié d'un espace de couleur pour la détection de la peau est un problème délicat [205]. Plusieurs de ces espaces ont été appliqués au problème de la modélisation de la couleur de la peau. Ce choix peut être guidé par deux observations:

- Etant donné que la couleur de la peau varie d'une personne à une autre, diverses études ont prouvé que la variation se situe plus au niveau de la composante de luminance qu'au niveau des composantes de chrominance [92] [116].

- Les pixels correspondant à la peau sont "assez bien" regroupés dans le plan des chrominances (voir la figure 1.15) [184].

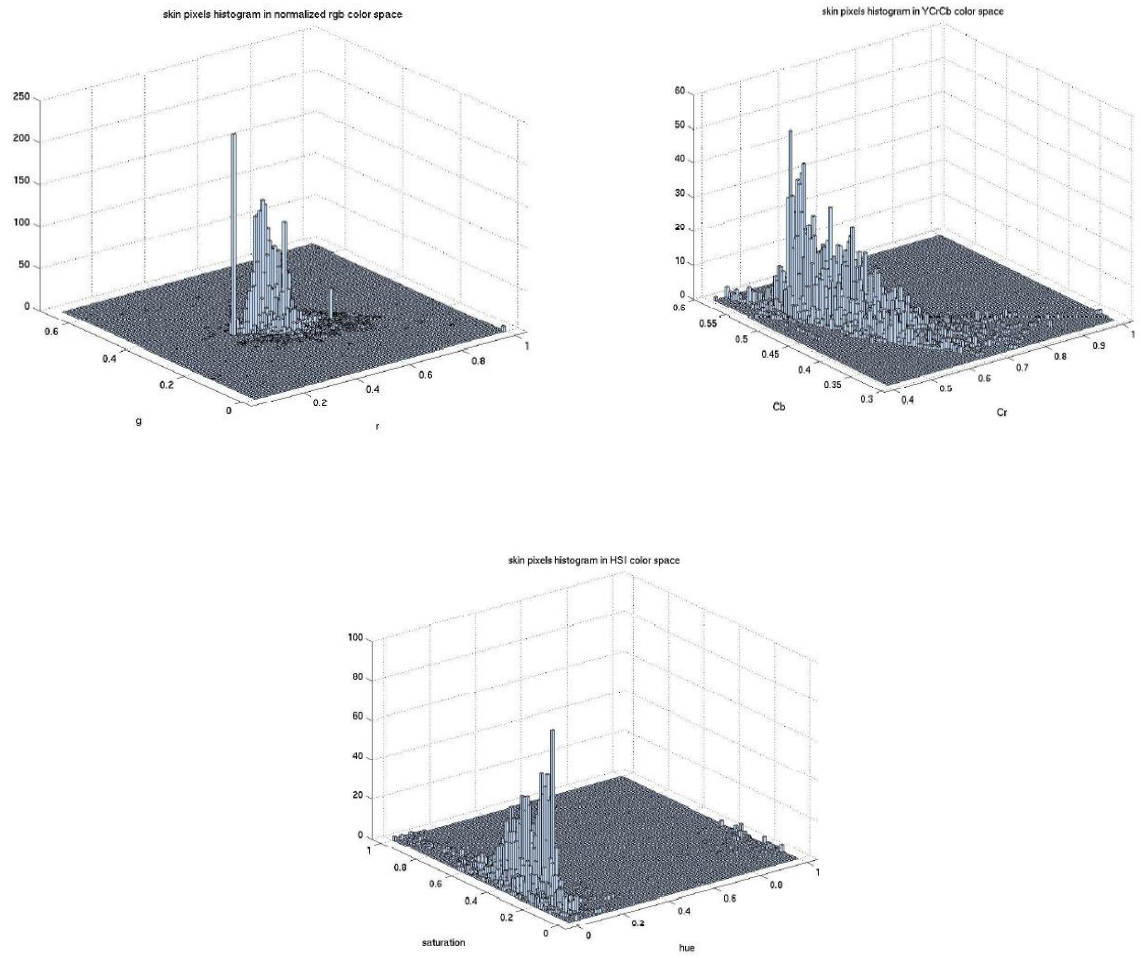


Figure. 1.15: Histogramme des pixels de peau dans le sous-espace *rg, CbCr et HS* [184].

Tenant compte de ces observations, les auteurs se sont penchés plus à l'étude des systèmes qui séparent les composantes de chrominance et de luminance afin de se focaliser exclusivement aux composantes de chrominance.

L'espace de couleur HSV (Hue, Saturation Value), ainsi que ses variantes, HSI (*Hue, Saturation, Intensity*) et HLS (*Hue, Lightness, Saturation*) ont été utilisés fréquemment dans beaucoup de travaux. D'une autre part, De très nombreux articles [165] [149] [107] [37] et [4] affirment que l'espace de couleur YCbCr permet une bonne détection des pixels de la peau. Certains articles affirment de leur coté qu'une bonne détection peut se faire dans d'autres espaces de couleur, tels que les espaces YIQ [26] [215], YUV [51] [147], XYZ de CIE [218] et CIE L*a*b* [33].
Terrillon et al. dans [199] ont effectué une étude comparative de 9 espaces de représentation. Ils ont montré que les meilleurs espaces de représentations des couleurs pour la détection de la peau sont les espaces normalisés *rgb* et *TSL*. L'espace *TSL* représente un espace de représentation perceptuel qui permet de

séparer les composantes de teinte, de saturation et de luminance. Il est semblable à l'espace HSI.

Nicolas Mottin [154] de son coté a effectué des tests comparatifs sur les meilleurs espaces de couleurs permettant de distinguer les pixels de peau de ceux de non peau. Il a choisi de tester 5 types d'espaces de couleurs: YUV, RGB, HSI, TSL, et Lab. L'auteur a exploité la base peau de Von Luschan illustrée sur la figure 1.16, qui contient 36 imagettes de peau pour effectuer ses tests. Cette base est organisée de manière à classer les peaux depuis les plus claires aux plus mates. Ces résultats ont montré que la meilleure détection de peau peut se faire dans l'espace HSI.

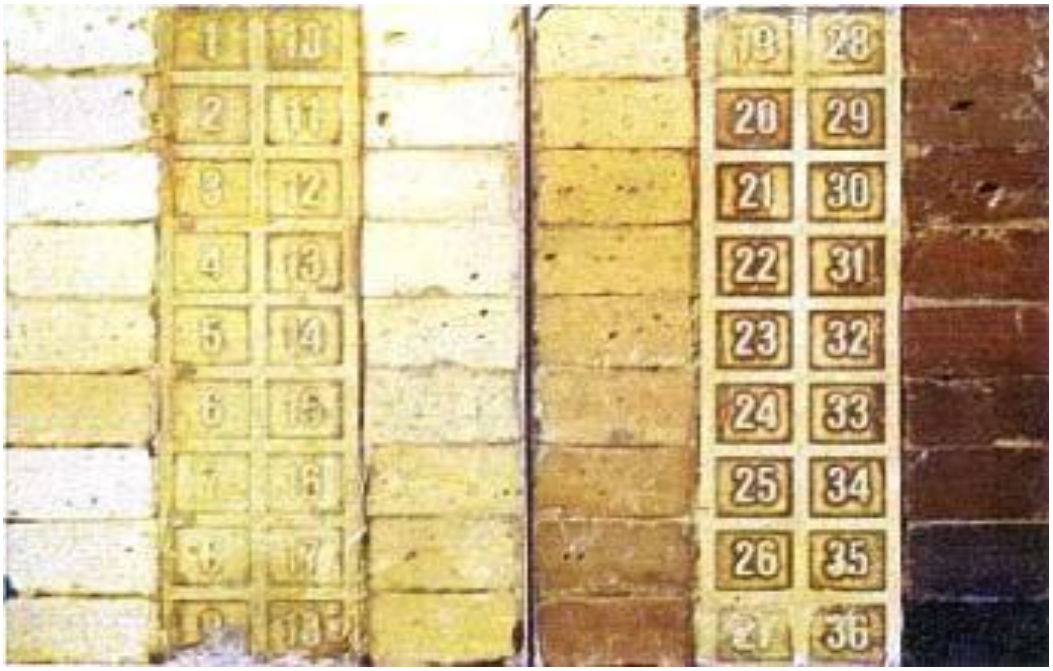


Figure. 1.16: Base de peaux de Von Luschan [84].

Girondel [84] a exploité lui aussi la base de Von Luschan afin d'effectuer ses test et a obtenu des résultats médiocres dans l'espace 3D RGB. Sur l'espace rg, il a utilisé 4 seuils (2 pour chaque composante normalisé) qui découlent de la distribution observée et a obtenu une distribution assez étalée des pixels peau. Les résultats obtenus étaient insatisfaisants. Selon lui, il est impossible de trouver des seuils ou une modélisation pour la distribution obtenue donnant des résultats concluants.

D'un autre coté, dans les travaux [155] [169] [26] [117], l'espace de couleur RGB a été utilisé pour la segmentation d'image de couleur de peau, mais l'espace de représentation le plus utilisé est l'espace normalisé *rgb* [222]. De nombreux articles recommandent une détection de peau dans cet espace [183] [8] [220] [226] [27] [189] [108] [221]. Cet espace est obtenu à partir de l'espace RGB par simple normalisation de la manière suivante:

$$r = \frac{R}{R+G+B} \quad g = \frac{G}{R+G+B} \quad b = \frac{B}{R+G+B}$$

Généralement, on se contente de la connaissance des deux premières composantes r et g puisqu'on a: r+g+b=1.

Dans l'article [221], des tests ont été effectués sur une base d'un millier de visages et ont donné suite à l'apparition de seuils qui donnent de bons résultats sur l'espace RGB. Ces seuils ont été obtenus d'une modélisation gaussienne de chaque composante. Les seuils obtenus pour l'ensemble de la base d'images sont:

$m_R = 234.29$ $\sigma_R = 26.77$
$m_G = 185.72$ $\sigma_G = 30.41$
$m_B = 151.11$ $\sigma_B = 25.68$

Cet article donne aussi (pour l'ensemble des visages de la base de données) des seuils pour l'espace rg:

$m_r = 104.22$ $\sigma_r = 4.93$
$m_g = 81.59$ $\sigma_g = 3.89$

En adaptant ces deux seuils, les résultats obtenus par [221] ne sont pas toujours satisfaisants. Dans certains cas, aucune zone de peau n'est détectée alors que sur d'autres, de grandes zones contenant ou non de la peau sont détectées comme peau.

Le tableau suivant illustre une étude comparative de certains espaces de couleurs à travers des travaux précédemment effectués et montre les avantages et inconvénients de l'utilisation de ces derniers [46].

Tableau. 1.1: Comparaison des différents espaces de couleur.

Esp.couleur	Référence	Méthode	Avantage	Inconvénient
RGB	[75] [84] [85] [221]	Explicitly Skin Color Boundary	1. Employé couramment pour stocker et traiter des images numériques. 2. Taux élevés de détection avec le modèle gaussien	1. Non fiable dû à la forte corrélation entre les composantes R, G et B.
	[116]	Statistical Model	d'histogramme et les modèles basés sur les	2. taux de faux positifs importants.
	[145]	Beta Mixture Models	distributions beta.	3. Inadéquat pour une bonne détection dans des conditions
	[26]	Bayes SPM		d'illumination variées. 4. Les régions peau

				et non peau peuvent avoir des valeurs colorimétriques similaires. 5. Echec de détection des parties peau ambrées. 6. Difficulté d'adaptation aux méthodes explicites.
rgb (RGB normalisé)	[153] [64] [27]	Neural network Expectation Maximization Algorithm Self Organized Map (SOM)	1. Ne contient pas de composante de luminance. 2. Vitesse de calcul améliorée du à l'utilisation de deux composantes uniquement. 3. Résultats satisfaisants dans les méthodes basées sur les distributions gaussiennes.	1. Performances diminuées dans des conditions d'illumination variées. 2. Perte de données dû à la non considération de l'information de luminance. 3. Echec de détection des parties peau ambrées.
Espaces de couleur perceptuels (HSV, HSI, HSL, TSL)	[185] [118] [199] [27]	Normalized Lookup Table(LUT) Explicitly defined skin region Gaussian model SOM	1. Robustesse aux changements d'illumination mineurs. 2. Bonne estimation des distributions de la couleur peau. 2. Les régions peau sont facilement identifiables dans les méthodes explicites. 3. Performance de détection élevée.	1. Conversion couteuse à partir de l'espace RGB. 2. Incapable de détecter des personnes appartenant à des races différentes dans les méthodes explicites. 3. Faux positifs provoqués par une illumination ou ombre intense sur les parties peau.
Espaces de couleur luminance/ chrominence YCbCr	 [133]	 Explicitly Defined Skin Region	1. Similaire à TSL en termes de séparation de la luminance et de la chrominance. 2. Largement utilisé dans beaucoup de travaux.	1. Echec de détection des parties peau ambrées. 2. Difficulté de détection dans des conditions de forte et faible illumination.

YCbCr YCbCr YIQ	[107] [26] [136]	Gaussian Model Explicitly defined skin region Single Gaussian Model Mixture Gaussian Model		3. Les régions peau et non peau peuvent avoir des valeurs colorimétriques similaires.

D'autres types d'espaces de couleurs ont été proposés afin d'améliorer la performance de détection. Parmi ces espaces, nous pouvons citer les espaces hybrides et les espaces normalisés. Les espaces de couleur hybrides sont issus de méthodes d'analyse et de sélection des composantes couleurs les plus discriminatives. La dimension d'un espace appartenant à cette famille n'est pas restreinte à trois composantes comme le sont les espaces de couleur classiques. Cela dépend de l'algorithme de sélection de ces composantes. Dans [207], les auteurs ont utilisé les espaces de couleurs RGB, rgb, XYZ, xyz, L*a*b*, L*u*v*, $I_1I_2I_3$, YIQ, YUV…, puis regroupent les composantes couleurs issues de ces systèmes de représentation de la couleur dans un espace couleur multidimensionnel (58 dimensions au total). Les auteurs effectuent une analyse globales sur l'ensemble des composantes et appliquent une procédure de sélection mesurant le pouvoir discriminant de chaque sous espace de représentation afin de déterminer l'espace de dimension préalablement inconnu le plus adéquat pour la détection.

Dans [223], les auteurs ont effectué leur test sur une base de 36818 images de visages et ont montré que les meilleurs espaces de couleur pour la détection sont ceux dont la somme de la deuxième et troisième ligne de la matrice de transformation à partir de l'espace RGB est égale à 0. Parmi les espaces qui connaissent cette propriété, les auteurs ont cité les espaces $I_1I_2I_3$, YUV, YIQ et LSLM. Les espaces de primaires RGB, XYZ et les espaces hybrides XGB, YRB et ZRG ne sont pas concernés par cette caractéristique et sont donc considérés comme espaces de couleurs inférieurs. Afin d'améliorer ces derniers les auteurs proposent des techniques de normalisation notée CSN (Color Space Normalization) opérant sur les matrices de transformation de telle manière à avoir la somme des éléments de la deuxième et la troisième ligne égale à 0.

L'une des techniques consiste à calculer d'abord la moyenne de la deuxième ligne notée m_2 ainsi que celle de la troisième ligne notée m_3 comme montré dans les équations suivantes:

$$\begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} = A \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1.23)$$

$m_2 = (a_{21} + a_{22} + a_{23})/3$ et $m_3 = (a_{31} + a_{32} + a_{33})/3$. L'espace de couleur normalisé est défini comme suit:

$$\begin{bmatrix} C_1' \\ C_2' \\ C_3' \end{bmatrix} = \tilde{A}\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} A_1 \\ \tilde{A}_2 \\ \tilde{A}_3 \end{bmatrix}\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} - m_2 & a_{22} - m_2 & a_{23} - m_2 \\ a_{31} - m_3 & a_{32} - m_3 & a_{33} - m_3 \end{bmatrix}\begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1.24)$$

Après normalisation, les auteurs dans [223] ont montré que les espaces RGB, XYZ, XGB, YRB et ZRG normalisés sont aussi efficaces pour la détection que les espaces $I_1I_2I_3$, YUV, YIQ et LSLM.

1.5 Conclusion

La détection de la peau a fait objet de plusieurs travaux de recherche. Elle se trouve au confluent de plusieurs disciplines comme la vidéosurveillance et le filtrage de sites web. Elle représente une étape cruciale indispensable aux seins de ces dernières. Il existe dans la littérature de nombreuses méthodes pour la détection de régions de peau dans une image. Trois approches se détachent principalement: la détection basée sur l'extraction de traits caractéristiques du visage et /ou des mains, la détection basée sur une approche couleur et la détection basée sur le mouvement.

Les approches basées sur l'extraction de traits caractéristiques reposent sur la définition de modèles pour ces traits. Le *template matching* qui consiste à comparer l'intensité de pixels entre un template prédéfini et plusieurs sous régions de l'image que l'on désire analyser est effectué. Les techniques basées sur le mouvement, telle que la méthode de différence d'image sont des techniques simples permettant de faire rapidement une estimation de la position d'un objet en mouvement. On trouve cette approche souvent utilisée conjointement avec l'information de couleur de peau ou couplée avec un système basé sur la reconnaissance de traits caractéristiques.

Nous avons opté pour une méthode basée sur la détection de peau par une approche couleur. Nous avons voulu que notre modèle de peau soit adapté pour les images couleurs fixes. De ce fait, Les approches basées sur le mouvement et l'extraction de traits caractéristiques ne correspondent pas à notre besoin. Nous avons alors élaboré une méthode basée sur une approche couleur. Il existe une multitude d'espaces de couleurs. Ceux-là sont regroupés en quatre familles: les systèmes de primaires, les systèmes luminance-chrominance, les systèmes perceptuels et les systèmes d'axes indépendants. La multitude des systèmes de représentation de la couleur pousse de nombreux chercheurs à déterminer quel est le système le mieux adapté face à un problème d'analyse d'images pour la détection de la peau. Les résultats des différentes études comparatives effectuées sur les différents espaces de couleurs ne sont pas décisifs. Chaque auteur privilégie un ou plusieurs espaces sur d'autres en se basant sur sa propre expérimentation. Hors, le choix d'un espace de couleur est tributaire de la méthode de classification adoptée et des données à tester. Il n'existe donc pas de meilleur espace de couleur. Seule une évaluation expérimentale comparative peut déterminer l'espace de représentation le plus approprié à la méthode utilisée et aux données de test.

Ce chapitre nous a fait prendre connaissance des notions fondamentales liées à la détection de la peau et aux espaces de couleur que nous utiliserons tout au long de notre travail. Dans le chapitre suivant, nous allons détailler les approches basées sur la couleur et étudier comment la couleur peut être exploitée dans le domaine de la détection de la peau.

CHAPITRE2
APPROCHES DE DETECTION DE LA PEAU BASEES SUR LA COULEUR

2.1 Introduction

L'information de couleur constitue un moyen simple et efficace pour la distinction entre les pixels peau et ceux de non peau. Il a été montré que la couleur de la peau est située dans un intervalle étroit de l'espace de couleur [77] [78] [124]. Cette information est donc facilement utilisable afin de discriminer les valeurs des pixels peau et ceux de non peau. La méthode généralement employée pour détecter la peau repose sur la modélisation de la distribution des vecteurs de chrominance dans un espace de représentation choisi. Les résultats obtenus dépendent surtout de l'espace et de la modélisation adoptés [199] [222]. La majorité des modèles de peau proposés se basent sur une méthode de segmentation basée sur le calcul d'histogramme de couleurs [212].

Dans le but d'avoir une bonne classification des pixels de peau de ceux de non peau, dans le cadre de l'approche couleur, il serait judicieux de tenir en compte non seulement d'une bonne modélisation permettant de distinguer les pixels de couleur peau de ceux de non peau, mais également le choix approprié de l'espace de couleur à adopter.

Nous allons présenter dans les sections suivantes les modèles de peau basés sur la couleur comme les techniques de classification paramétriques et les techniques de classification non paramétriques, mais avant de dénombrer ces modèles de peau et recenser les méthodes les plus utilisées dans chaque modèle, nous devons nous focaliser sur la base théorique soutenant ces méthodes afin de faciliter la compréhension de la majorité des modèles de peau présentées dans les sections suivantes.

2.2 Fondements théoriques

Il existe deux grandes familles de méthodes de classification: le mode supervisé et le mode non supervisé. On parle de mode supervisé lorsque l'on dispose d'un ensemble de points étiquetés pour l'apprentissage. On considère dans ce cas qu'il existe déjà une classification préétablie. Cela suppose que les classes sont donc déjà connues et chaque classe lui est attribuée des exemples préalablement. Sinon, nous parlons d'une classification non supervisée ou automatique. Dans le cas de la détection de la peau, les classes de peau et de non peau doivent se connaitre à l'avance. On applique alors dans notre cas les méthodes de classification supervisées. On dispose donc d'une connaissance à priori de l'appartenance de chaque échantillon de l'ensemble d'apprentissage à une classe donnée, ce qui revient à supposer une connaissance à priori sur l'image à segmenter. A partir des

données expérimentales que forment les échantillons tirés des classes, on obtient un effet d'apprentissage. Le système s'organise en vue de discriminer les échantillons ultérieurs à partir des échantillons qui lui sont soumis. On dit que le système est capable de généraliser à partir des échantillons d'apprentissage.
Deux catégories de méthodes de classification sont à distinguer:

- les méthodes indirectes qui utilisent la formule de Bayes.

- les méthodes directes qui évaluent les probabilités à posteriori sans faire intervenir la formule de Bayes.

La formule de Bayes permet de déterminer les probabilités d'appartenance à posteriori si les densités de probabilité et les probabilités à priori sont connues. La règle de Bayes permet d'obtenir le taux d'erreur de classification minimum, ce qui est l'objectif souhaitable pour tout système de classification.
La majorité des travaux de recherche sur la modélisation de la peau se basent sur les méthodes indirectes. Il existe deux types de méthodes indirectes:

- les méthodes paramétriques: font usage d'une hypothèse sur la forme analytique de la distribution.

- les méthodes non paramétriques: ne font usage d'aucune hypothèse sur la forme de distribution.

La conception d'un système de classification demande:

- La spécification des classes {$Classe_k$}.

- La sélection des caractéristiques $\vec{X}$ des classes.

- La spécification d'une représentation (modèle) pour $P(\vec{X}/Classe_k)$ et $P(Classe_k)$. Cette spécification peut être une fonction paramétrique ou non-paramétrique.

- L'estimation des paramètres $P(\vec{X}/Classe_k)$ à partir d'un ensemble $S_k = \{\vec{X}_m\}$ de M_k observations.

Dans notre cas, pour chaque observation $\vec{X}$, la classe $Classe_k$ est connue, et peut être la classe peau ou la classe non-peau. Il suffit donc d'estimer les densités de probabilité [96].

2.2.1 Règle de décision de Bayes

La classification en utilisant la règle de Bayes se résume en une division de l'espace des caractéristiques en partitions disjointes. L'expression mathématique de la formule de Bayes prend en considération la probabilité à priori d'apparition des individus des différentes classes et de leur distribution dans l'espace des descripteurs:

$$P(Classe_k / \vec{X}) = \frac{P(\vec{X} / Classe_k)\ p(Classe_k)}{p(\vec{X})} \qquad (2.1)$$

avec:

$P(Classe_k / \vec{X})$: probabilité à posteriori qu'un pixel de caractéristiques $\vec{X}$ appartienne à la
classe k.

$P(\vec{X} / Classe_k)$: densité de la probabilité $\vec{X}$ si la classe est k.

P ($Classe_k$): probabilité à priori qu'un pixel appartienne à la classe k.
P($\vec{X}$): probabilité d'observer $\vec{X}$.

La règle de décision de Bayes consiste à affecter l'individu à la classe dont la probabilité à posteriori (calculée par la formule de Bayes ou par toute autre méthode) est la plus grande.
Cette décision minimise la probabilité d'erreur de classement. La figure 2.1 [191] apporte une explication géométrique.

Les courbes correspondent aux fonctions de densités de probabilité pondérées par les probabilités à priori correspondant aux deux classes A et B. De cette manière, ces courbes sont directement reliées à la densité des individus. Le trait vertical marque le seuil de classification donné par la règle de Bayes entre les deux classes.

L'erreur de classification est le nombre d'exemples A classée comme B et inversement. Elle correspond à la surface d'intersection des deux courbes. Si l'on choisit un autre seuil, nous nous apercevons que la nouvelle erreur de classification est égale à l'erreur de classification de Bayes augmentée d'une contribution positive. Elle est donc toujours supérieure à l'erreur de Bayes. Ainsi quel que soit le seuil pris pour séparer les 2 classes, l'erreur de classification est toujours supérieure à celle trouvée avec la règle de Bayes [96].

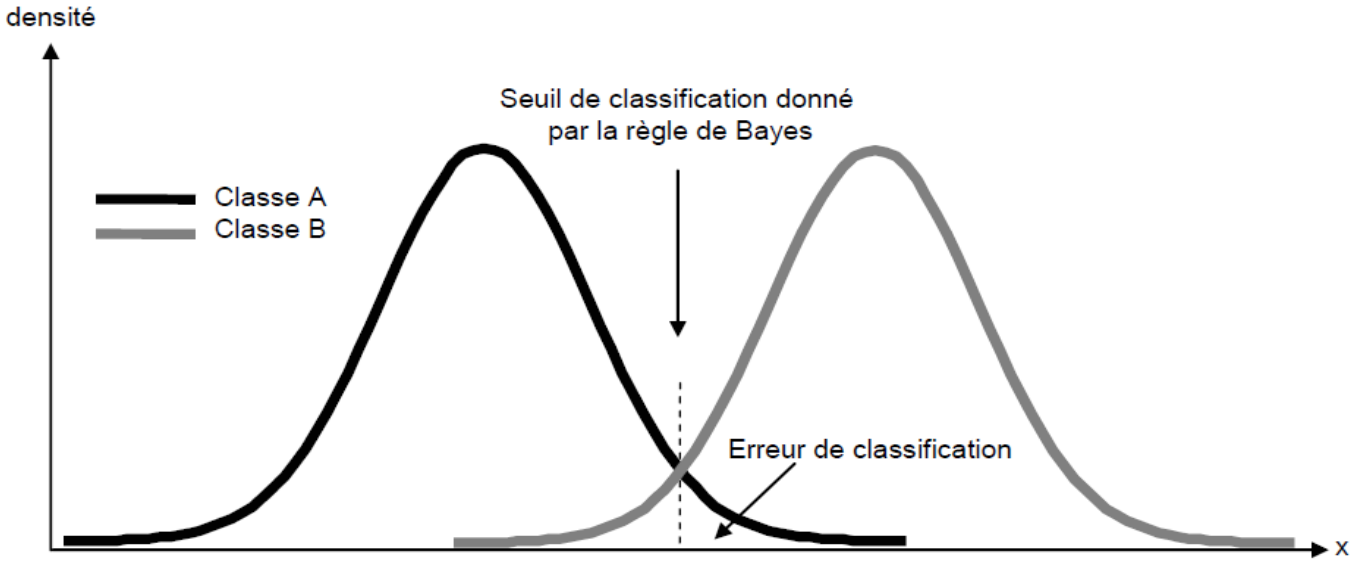


Figure. 2.1: Distribution des individus dans l'espace de description.

2.2.2 Estimation paramétrique des densités de probabilité

Les méthodes paramétriques consistent à faire une hypothèse concernant la forme analytique de la distribution de probabilité recherchée, et à estimer les paramètres de cette distribution à partir des données dont on dispose. Il s'agit d'ajuster une loi de distribution choisie par rapport aux individus étudiés à l'aide de quelques paramètres comme la moyenne ou la variance. En se basant sur

l'estimation des paramètres, on peut utiliser la forme analytique de la densité ainsi déterminée pour en déduire la densité en tout point de l'espace de représentation.

L'hypothèse la plus répondue est que la répartition des individus de chacune des classes suit une loi gaussienne (Figure 2.2 [96]). Cette distribution "*normale*" des individus est la plus utilisée et conduit à la méthode appelée "analyse discriminante avec une règle d'affectation probabiliste" [96].

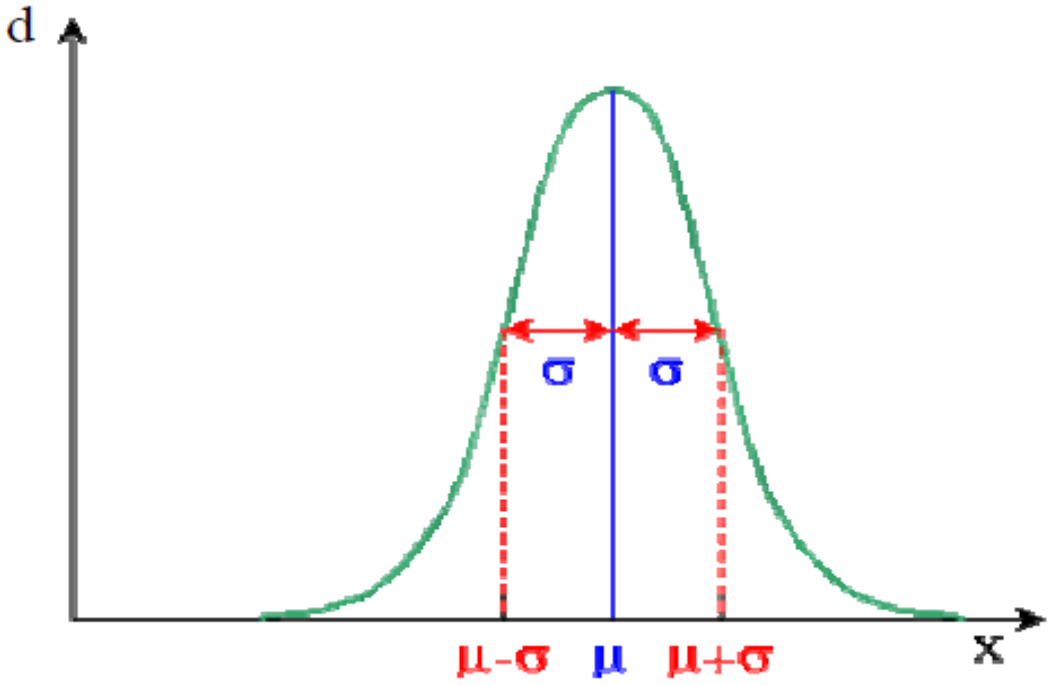


Figure. 2.2: Distribution relative à une loi normale.

μ: moyenne de la gaussienne.
σ: écart type
d: distribution des individus

2.2.2.1 Analyse discriminante avec une règle d'affectation probabiliste

L'hypothèse de distribution consiste en une loi gaussienne multidimensionnelle qui se présente comme suit:

$$P(x/Classe_k) = N(\vec{\mu}, \sum) = \frac{1}{(2\Pi)^{n/2}\sqrt{|\sum_k|}} e^{-\frac{1}{2}(x-\mu_k)^T \sum_k^{-1}(x-\mu_k)} \quad (2.2)$$

avec:

$\sum_k$: Matrice de covariance de la classe k.

μ_k : Moyenne de la gaussienne de la classe k.

La matrice de covariance et le centre de la classe k sont estimés par la matrice de covariance et la moyenne des individus appartenant à la classe k. A partir des estimations des matrices de covariance, des moyennes des gaussiennes (pour chacune des classes) et des probabilités à priori, on calcule par la formule de Bayes les probabilités à posteriori d'appartenance aux classes. La règle de décision consiste à classifier un individu dans la classe qui obtient la plus grande probabilité à posteriori. La frontière de séparation est donc déterminée par l'ensemble des points pour lesquels les probabilités à posteriori sont égales [191].

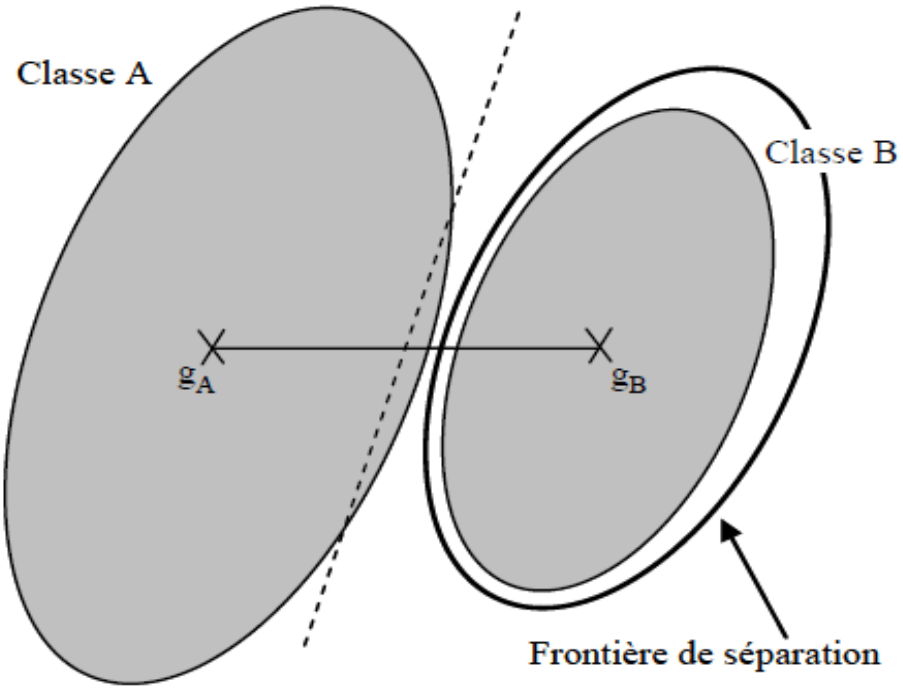


Figure. 2.3: Frontière de séparation [191].

On remarque sur la figure 2.3, que la frontière de séparation prend en considération la différence des classes. Le trait pointillé matérialise la frontière obtenue avec la règle d'affectation géométrique qu'on présentera dans la section suivante.

Un inconvénient majeur de cette méthode est que les estimations des différentes matrices sont effectuées à partir des exemples qui risquent d'être peu nombreux. Elles sont donc très sensibles aux exemples marginaux. Il est préférable alors d'ajouter quelques hypothèses qui conduisent à l'analyse discriminante avec une règle d'affectation géométrique [191].

2.2.2.2 Analyse discriminante avec une règle d'affectation géométrique

Il s'agit de la forme la plus simple de l'analyse discriminante. Afin de converger vers la règle de Bayes, L'hypothèse de départ est complétée par les hypothèses suivantes:

- Les individus d'une classe k sont répartis suivant une loi gaussienne multidimensionnelle déterminée par l'équation 2.2.

- Les différentes matrices de covariance sont identiques: $\sum_1 = \sum_2 = \ldots. = \sum_c = \sum$

- Les probabilités à priori des classes sont identiques: $Pr_1 = Pr_2 = \ldots Pr_c = 1/C$

Cette méthode se base sur la métrique de Mahalanobis (notée $\Delta_\sum$) pour classer un nouvel exemple. Elle affecte à cet exemple la classe correspondant à la plus petite distance le séparant avec les centres de gravité des autres classes et est définie dans l'espace de description par la matrice de covariance des individus:

$$\Delta^2_\sum(\mu_1,\mu_2) = (\mu_1 - \mu_2)^T \sum^{-1} (\mu_1 - \mu_2) \quad (2.3)$$

avec μ_1 μ_2 : 2 vecteurs dans l'espace de description.

Cette méthode est dite géométrique car elle ne tient compte que de l'éloignement de l'exemple considéré aux centres de gravité. Dans le cas de la classification à deux classes, on introduit la fonction discriminante de Fisher [75] qui est donnée par:

$$w = e^T \sum^{-1} (\mu_1 - \mu_2) \quad (2.4)$$

où w est la valeur de la fonction discriminante de Fisher au point de coordonnées e et μ_k est la moyenne de la gaussienne de la classe k. Ainsi, on affectera l'observation e à la classe 1 si:

$$w = e^T \sum^{-1} (\mu_1 - \mu_2) > \frac{1}{2} (\mu_1 - \mu_2)^T \sum^{-1} (\mu_1 - \mu_2) \quad (2.5)$$

La figure suivante [191] montre un exemple de classification à deux classes:

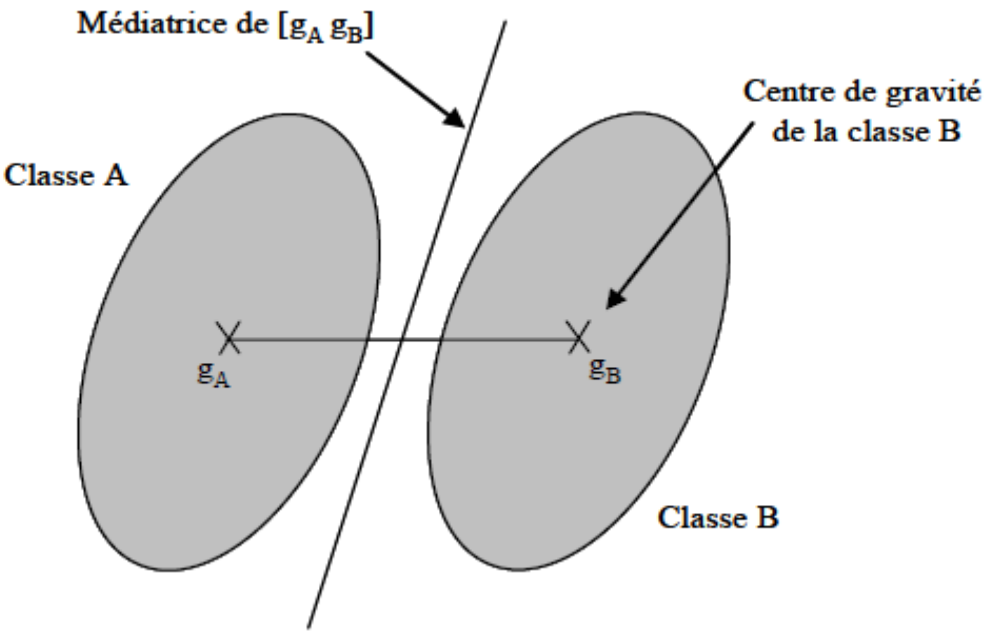


Figure. 2.4: Frontière de séparation (Analyse discriminante avec une règle d'affectation géométrique).

Selon la métrique de Mahalanobis, la frontière entre les deux classes (A et B) est bien la médiatrice du segment [g_A,g_B]. Malheureusement le résultat obtenu est rarement celui que l'on obtiendrait par le classifieur de Bayes. Ainsi, une configuration typique est celle de la figure suivante [191]:

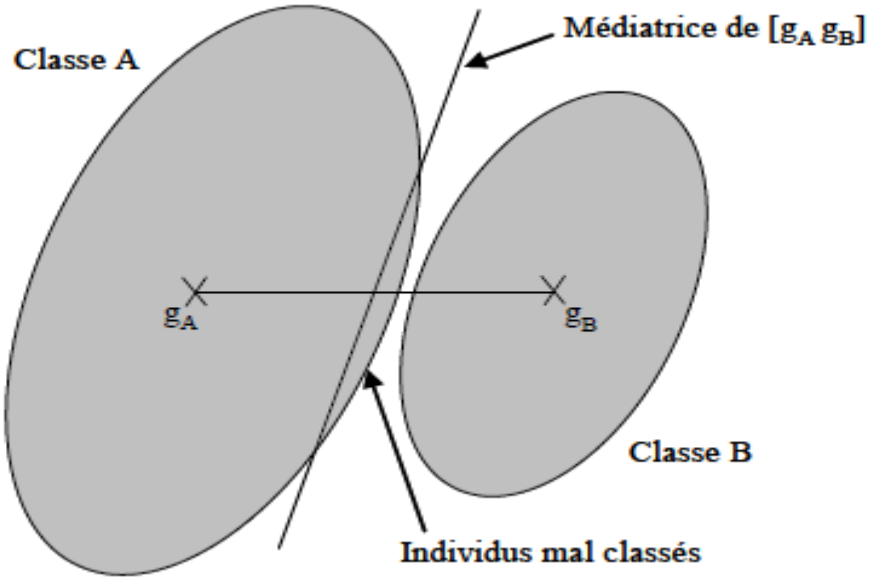


Figure. 2.5: Individus mal classés (analyse discriminante avec une règle d'affectation géométrique).

Les individus de la classe A sont plus dispersés que ceux de la classe B. La frontière, quant à elle est restée la même puisque les centres de gravité sont restés identiques. De nombreux individus de la classe A sont donc mal classés [96].

2.2.2.3 Estimation de la moyenne et de la covariance

Les densités de probabilité de $\vec{X}$ peuvent être représentées par une loi normale comme suit [96]:

$$P(\vec{X}/Classe_k) = N(\vec{\mu}, \sum) = \frac{1}{(2\pi)^{n/2}\sqrt{|\sum_k|}} e^{-\frac{1}{2}(\vec{X}-\vec{\mu})^T \sum_k^{-1}(\vec{X}-\vec{\mu})} \quad (2.6)$$

Il est nécessaire dans ce cas de pouvoir estimer la moyenne et la covariance.

2.2.2.3.1 Estimation de la moyenne [96]

La moyenne μ_i représente l'espérance de {x_i}. Elle est donnée par la formule:

$$\mu \equiv E\{x\} = \sum x\, p(x) \quad (2.7)$$

Pour les vecteurs de propriétés:

$$\vec{\mu} \equiv E\{\vec{X}\} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_n \end{pmatrix} = \begin{pmatrix} E\{x_1\} \\ E\{x_2\} \\ \dots \\ E\{x_n\} \end{pmatrix} \quad (2.8)$$

Chaque composante se présente comme suit:

$$\mu_i \equiv E\{x_i\} = \sum x_i\, p(x_i) \quad (2.9)$$

Cette moyenne peut être estimée par l'histogramme. Soit M le nombre d'individus sur lesquelles s'effectuent les observations.

$$\Pr(X = x) = P(x) = \frac{1}{M} h(x)$$

(2.10)

On a donc:

$$\mu = E\{x\} = \sum_{x_{\min}}^{x_{\max}} P(x).x \approx \frac{1}{M} \sum_{x_{\min}}^{x_{\max}} h(x).x \quad (2.11)$$

Pour un vecteur de variables à n dimensions, on aura:

$$\vec{\mu} = E\left\{\vec{X}_n\right\} = \sum_{x_{\min}}^{x_{\max}} P(\vec{X}_n).\vec{X}_n \approx \frac{1}{M}\sum_{x_{\min}}^{x_{\max}} h(\vec{X}_n).\vec{X}_n \qquad (2.12)$$

2.2.2.3.2 Estimation de la covariance [96]

Le deuxième moment de la densité de probabilité consiste en la variance notée σ^2. Soit M le nombre d'individus sur lesquelles s'effectuent les observations.

$$\sigma^2 = E\left\{(x-\mu)^2\right\} = \frac{1}{M}\sum_{i=1}^{M}(x_i-\mu)^2 \qquad (2.13)$$

Mais l'usage de μ estimé avec le même ensemble, introduit un biais dans σ^2. Pour l'éviter, on utilise une estimation sans biais.

$$\sigma^2 = \frac{1}{M-1}\sum_{i=1}^{M}(x_i-\mu)^2 \qquad (2.14)$$

Avec μ et σ^2 , on peut estimer la densité p(x) par:

$$P(x) = N(\vec{\mu},\sigma) = \frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (2.15)$$

σ^2 est le deuxième moment de P(x):

$$\sigma^2 = E\left\{(x-\mu)^2\right\} = \sum_{x=x_{\min}}^{x_{\max}} P(x).(x-\mu)^2 = \frac{1}{M}\sum_{x=x_{\min}}^{x_{\max}} h(x).(x-\mu)^2 \qquad (2.16)$$

Pour N dimensions, la covariance entre les variables x_i et x_j est estimée comme suit:

$$\sigma_{ij}^2 = E\left\{(x_i - E\{x_i\})(x_j - E\{x_j\})\right\} \quad \text{donc :} \quad \sigma_{ij}^2 = \frac{1}{M}\sum_{m=1}^{M}(x_{im}-\mu_i)(x_{jm}-\mu_j).$$

et encore pour éviter le biais, on utilise:

$$\sigma_{ij}^2 = \frac{1}{M-1}\sum_{m=1}^{M}(x_{im}-\mu_i)(x_{jm}-\mu_j) \qquad (2.17)$$

Ces coefficients composent la matrice de covariance Σ:

$$\sum_x \equiv E\left\{(\vec{X}-\vec{\mu})(\vec{X}-\vec{\mu})^T\right\} = E((\vec{X}-E\{\vec{X}\})(\vec{X}-E\{\vec{X}\})^T) \qquad (2.18)$$

$$\sum_x \equiv \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \dots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \dots & \sigma_{2n}^2 \\ \dots & \dots & \dots & \dots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \dots & \sigma_{nn}^2 \end{pmatrix} \quad (2.19)$$

On obtient donc les paramètres de la fonction:

$$P(\vec{X}/Classe_k) = N(\vec{\mu}, \sum) = \frac{1}{(2\pi)^{n/2}\sqrt{|\sum_k|}} e^{-\frac{1}{2}(\vec{X}-\vec{\mu})^T \sum_k^{-1}(\vec{X}-\vec{\mu})} \quad (2.20)$$

2.2.3 Estimation non paramétrique des densités de probabilité

Lorsqu'on ne peut pas faire d'hypothèse sur la distribution des individus, il faut se tourner vers des méthodes non paramétriques. Il s'agit de s'affranchir de l'hypothèse d'une loi paramétrique et d'estimer la distribution *P(x)= P(X=x/ K=i)* de la classe courante par approximation.

L'estimation non paramétrique de la densité de probabilité consiste à délimiter une région R_N autour d'un point considéré et de compter le nombre d'individus dans ce volume, et enfin de déterminer la densité comme le rapport entre ce nombre (divisé par le nombre total d'individus) et le volume de la région ([159], [67] et [18]). Ainsi, on obtient une estimation de la densité de probabilité avec la formule suivante:

$$\hat{P}_N(x) = \frac{K_N}{N.V_N} \quad (2.21)$$

avec:

N: nombre d'individus de l'échantillon
K_N: nombre d'individus dans la région R_N
V_N: volume de R_N

Ayant une telle fonction, on peut utiliser une technique de classification en appliquant la règle de Bayes. L'estimation de la densité de probabilité peut se faire par une table de fréquence d'occurrence (un histogramme).
Le problème principal dans l'estimation d'une fonction de densité pour un vecteur de caractéristiques est la croissance exponentielle du nombre de cellules N avec le nombre de dimensions D. Ceci induit une croissance exponentielle dans les nombres d'exemples M nécessaires. En Anglais, on appelle ce problème "The Curse of Dimensionality" [191].

2.3 Méthodes de détection de la peau

2.3.1 Les méthodes explicites

Les méthodes explicites utilisent des règles de décision empiriques et/ou statistiques [212] pour la détection des pixels ayant la couleur de la peau. Les images doivent être acquises dans un environnement contrôlé (i.e. un éclairage réglable) [221].
Une méthode explicite est une méthode de classification qui consiste à définir explicitement les frontières de la région peau dans l'espace de couleur utilisé. A titre d'exemple, Peer et al. [161] considèrent dans l'espace RGB un pixel comme pixel de couleur peau si chacune des conditions suivantes est respectée:

$$\begin{cases} R > 95 \text{ et } G > 40 \text{ et } B > 20 \text{ et} \\ \max\{R,G,B\} - \min\{R,G,B\} > 15 \text{ et} \\ |R - G| > 15 \text{ et } R > G \text{ et } R > B \end{cases} \quad (2.22)$$

Par ailleurs, Chai et Ngan [36] ont proposé un algorithme de segmentation de visage dans lequel ils ont employé les deux axes Cb et Cr du modèle couleur *YCbCr* pour déterminer les régions ayant la couleur peau. Ils ont utilisé la base de données de visages ECU [70]. Ils ont trouvé que les valeurs de pixels dans les domaines DCb = [77, 127] et DCr = [133, 173] définissent bien les pixels peau. De même, Garcia et Tziritas [82] ont segmenté la peau en utilisant les espaces YCbCr et HSV.
Chiunhsiun Lin et al. [139] considèrent qu'un nouveau pixel à traiter dans l'espace rgb est un pixel peau si les conditions suivantes sont vérifiées:

$$\begin{cases} r > \alpha \\ \beta_1 < (r-g) < \beta_2 \\ \lambda_1 < (r-b) < \lambda_2 \end{cases} \quad (2.23)$$

Les valeurs $\alpha = 80$, $\beta_1 = 0$, $\beta_2 = 56$, $\lambda_1 = 0$ et $\lambda_2 = 98$ sont ajustés expérimentalement.
Les règles proposées par Kukharev [132] et al. sur l'espace YCbCr pour identifier les pixels peau sont les suivantes:

$$\begin{cases} Y > 80 \\ 85 < Cb < 135 \\ 135 < Cr < 180 \end{cases} \quad (2.24)$$

où Y, Cb, Cr = [0, 255].

Afin de définir des intervalles de définition de la couleur peau dans les espaces YCbCr et HSV, J.A. Marcial-Basilio ont analysé des histogrammes sur plusieurs images de personnes de différentes race [148]. Ils ont constaté que dans l'espace YCbCr, la composante de luminance Y diffère considérablement d'une race à une autre. Ils se sont donc intéressés exclusivement aux composantes Cb et Cr et ont défini les intervalles suivant:

$$\begin{cases} 80 \leq Cb \leq 120 \\ 133 \leq Cr \leq 173 \end{cases} \quad (2.25)$$

Dans l'espace HSV, ils ont tenu compte aussi bien des composantes de chrominance que celle de luminance et ont défini les intervalles suivant:

$$\begin{cases} 0 < H < 0.25 \\ 0.15 < S < 0.9 \\ 0.2 < V < 0.95 \end{cases} \quad (2.26)$$

où H, S et V= [0, 1].

L'avantage de ces méthodes réside dans la simplicité des règles de détection de la peau qu'elles utilisent, ce qui permet une classification rapide. Cependant, leur problème principal est la difficulté de déterminer empiriquement un espace de couleur approprié ainsi que des règles de décision adéquates qui assurent un taux de reconnaissance élevé.

Une méthode utilisant des algorithmes d'apprentissage a été proposée pour résoudre ces problèmes [88]. Les auteurs commencent par choisir un espace RGB normalisé (où
$r = R/R+B+G$, $g = G/R+B+G$ et $b = B/R+B+G$ et la somme des trois composantes normalisées $r+g+b= 1$) sur lequel ils appliquent un algorithme d'induction constructive afin de créer de nouveaux ensembles d'attributs autre que les composantes rgb. Une règle de décision, semblable à l'équation (2.22) qui réalise la meilleure identification possible, est estimée pour chaque ensemble d'attributs. Ils ont obtenu des résultats meilleurs que ceux qui sont obtenus avec un classifieur de Bayes défini dans l'espace RGB [2].

Dans [195], les auteurs ont proposé une approche qui consiste à transformer l'image initiale à tester de l'espace RGB à l'espace YIQ. Ensuite, ils classent les pixels de l'image selon leurs valeurs dans le canal I de l'espace YIQ. Un pixel est affecté à la classe peau si et seulement si sa valeur dans l'espace YIQ se situe entre 20 et 90.

M. Abdullah-Al-Wadud et al. [4] ont proposé une méthode explicite qui se base sur la carte de distance de couleurs dans l'espace RGB. Ils définissent un centroide pour la classe peau notée SSC (Rs,Gs,Bs) (*standard skin color*) avec Rs, Gs et Bs les coordonnées de ce point dans l'espace RGB.

Si les intervalles de définition des composante R, G et B pour les pixels peau sont respectivement [R_{start}, R_{end}], [G_{start}, G_{end}], [B_{start}, B_{end}] alors le SSC correspondant est défini selon les coordonnées ((R_{start} + R_{end})/2, (G_{start} + G_{end})/2, (B_{start} + B_{end})/2). Ils définissent la distance colorimétrique comme la distance euclidienne entre un pixel donné et le centroide dans l'espace RGB normalisée par un scalaire défini comme le rayon d'un cercle dont l'origine est le SSC comme illustré sur la figure suivante [4]:

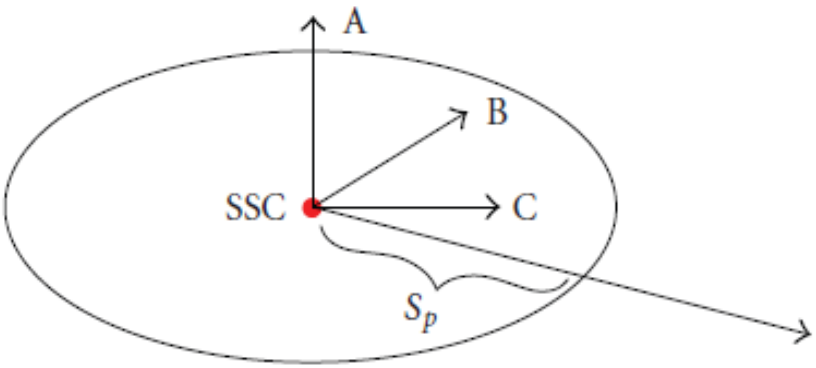


Figure. 2.6: Définition d'un rayon de normalisation pour le centroide.

La distance de couleur CD entre un point (R, G, B) et le SSC (R_s, G_s, B_s) est définie comme suit:

$$CD = \frac{\sqrt{(R-R_s)^2+(G-G_s)^2+(B-B_s)^2}}{SP} \tag{2.27}$$

La valeur de CD est considérée comme le potentiel d'une couleur d'être considérée comme couleur de peau. Plus celle-ci est petite, plus le pixel traité est considérée comme pixel de peau. La carte de distance de couleurs DM (Distance Map) est une image en niveau de gris définie comme suit:

$$DM(x,y) = \frac{d(x,y) - \min_{\forall x,y}(d(x,y))}{\max_{\forall x,y}(d(x,y)) - \min_{\forall x,y}(d(x,y))} \times 255 \tag{2.28}$$

où d(x,y) dénote la distance de couleur CD du point localisé par les coordonnées (x,y) de l'image traitée. Cependant, dans certains cas où il existe plus d'un cluster représentant les pixels peau, un centroide est défini pour chaque cluster et une DM initiale est créée pour chacun d'eux. La DM finale est générée en tenant compte pour chaque pixel de la CD ayant la plus petite valeur. En d'autres termes, la CD finale d'un pixel représente le minimum de ses CD dans les DM initiales. Les auteurs ont

employé les règles employés dans [83] et [129] pour la détection de la peau définies comme suit:

Sous l'illumination uniforme du jour:

$$\begin{cases} R > 95,\ G > 40,\ B > 20 \\ Max(R,G,B) - Min(R,G,B) > 15 \\ |R - G| > 15,\quad R > G,\quad R > B \end{cases} \tag{2.29}$$

Sous l'illumination latérale du jour ou d'une lampe:

$$\begin{cases} R > 220, G > 210, B > 170 \\ |R - G| \le 15, B < R, B > G \end{cases} \tag{2.30}$$

Un centroide est fixé au centre du cluster de peau désigné par la première ligne de (2.29) dont les coordonnées sont (R_s, G_s, B_s) = (175, 145.5, 137.5). Ensuite, une DM est calculé pour l'ensemble des pixels traités. Une valeur de 255 est attribuée aux pixels dont les valeurs colorimétriques ne satisfont pas les inégalités de la deuxième et troisième ligne de (2.29). Nous obtenons donc une image en niveau de gris Map_1. En utilisant la même stratégie, une autre image en niveaux de gris Map_2 est générée à partir de (2.30). Par la suite, les deux images sont combinées en une seule selon l'équation suivante:

$$M(x, y) = \text{Min}\{Map_1(x,y), Map_2(x,y)\} \tag{2.31}$$

où (x, y) représente les coordonnées des pixels dans l'image. Les inégalités dans (2.29) et (2.30) représentent deux clusters de peau dans l'espace RGB, tandis que l'équation (2.31) assure la représentation de la distance de chaque pixel au cluster le plus proche.

2.3.2 Modèles de peau paramétriques

Les approches paramétriques estiment la distribution de la couleur de la peau sous forme de modèles explicites. La distribution est généralement caractérisée par une densité ou une somme de densités de probabilité dont les paramètres sont obtenus à partir d'un ensemble d'apprentissage [222] [95].

Les pixels correspondants à la peau sont assez bien regroupés dans le plan des chrominances. On peut donc représenter la distribution de la couleur de la peau dans ce plan sous la forme d'une densité de probabilité en s'appuyant sur quelques fonctions spécifiques paramétriques. Les modèles les plus couramment utilisés sont le modèle gaussien qui représente cette distribution sous la forme d'une gaussienne [199] [177] et le modèle de mélange de gaussiennes (Mixture of Gaussians) qui représente la distribution sous la forme d'une somme de gaussiennes. Ce dernier

modèle est semble-t-il plus adapté pour prendre en compte la variabilité des conditions d'acquisition des images et la présence de populations hétérogènes [95] [32]. Nous présentons dans la suite les méthodes paramétriques les plus utilisées pour construire un modèle de peau et de non peau.

2.3.2.1 Modèle basé sur une simple gaussienne

La distribution de couleur de peau est estimée par une fonction de densité de probabilité gaussienne comme suit:

$$P(c/peau)=\frac{1}{(2\Pi)\sqrt{\left|\sum_{peau}\right|}}e^{-\frac{1}{2}(c-\mu_{peau})^{T}\sum_{peau}^{-1}(c-\mu_{peau})} \tag{2.32}$$

Où c est la variable aléatoire à deux dimensions représentant le couple de chrominance; μ_{peau} et Σ_{peau} sont respectivement l'espérance et la matrice de covariance représentant les paramètres du modèle gaussien. Ces paramètres peuvent être estimés à partir de l'échantillon d'apprentissage selon les équations suivantes [191]:

$$\mu_{peau}=\frac{1}{N_{peau}}\sum_{c\in C}N_{peau}(c)c \tag{2.33}$$

$$\sum_{peau}=\frac{1}{N_{peau}-1}\sum_{c\in C}N_{peau}(c)(c-\mu_{peau})(c-\mu_{peau})^{T} \tag{2.34}$$

avec: N_{peau} correspond au nombre de pixels peau et $N_{peau}(c)$ la fréquence d'apparition du pixel c dans la classe peau.

2.3.2.2 Modèle basé sur une mixture de gaussiennes

L'utilisation d'un modèle mixte de gaussiennes (*Gaussian Mixture Model*) a également été proposée par d'autres auteurs [115], [199], [197], [147], [108] et [221] afin de mieux représenter et modéliser la portion d'un espace de couleur associée à la couleur de peau. Le mélange des Gaussiennes est une extension des gaussiennes simples. Il a la capacité de représenter les distributions les plus complexes qu'une gaussienne simple. La fonction de densité de probabilité gaussienne est représentée comme suit [96]:

$$P(c/peau)=\sum_{n=1}^{N}w_n p_n(c/peau) \tag{2.35}$$

avec:

P_n: les noyaux des gaussiens définis dans (2.32). Chaque P_n est lui-même une distribution gaussienne.

N: le nombre de noyaux gaussiens qu'il faut choisir correctement de sorte que le modèle puisse bien représenter les données d'apprentissage

w_n: les poids des noyaux correspondants dont la somme est égale à 1.

Dans ce cas, trois paramètres sont à estimer (w_n, μ_n, Σ_n). Généralement, l'algorithme EM (*Expectation Maximization*) est appliqué pour estimer ces paramètres [96]. La figure suivante illustre un mélange de trois gaussiennes en proportions égales (à gauche) et l'histogramme correspondant (à droite). La courbe en trait plein à gauche représente la densité de probabilité théorique du mélange résultant en tenant compte des proportions.

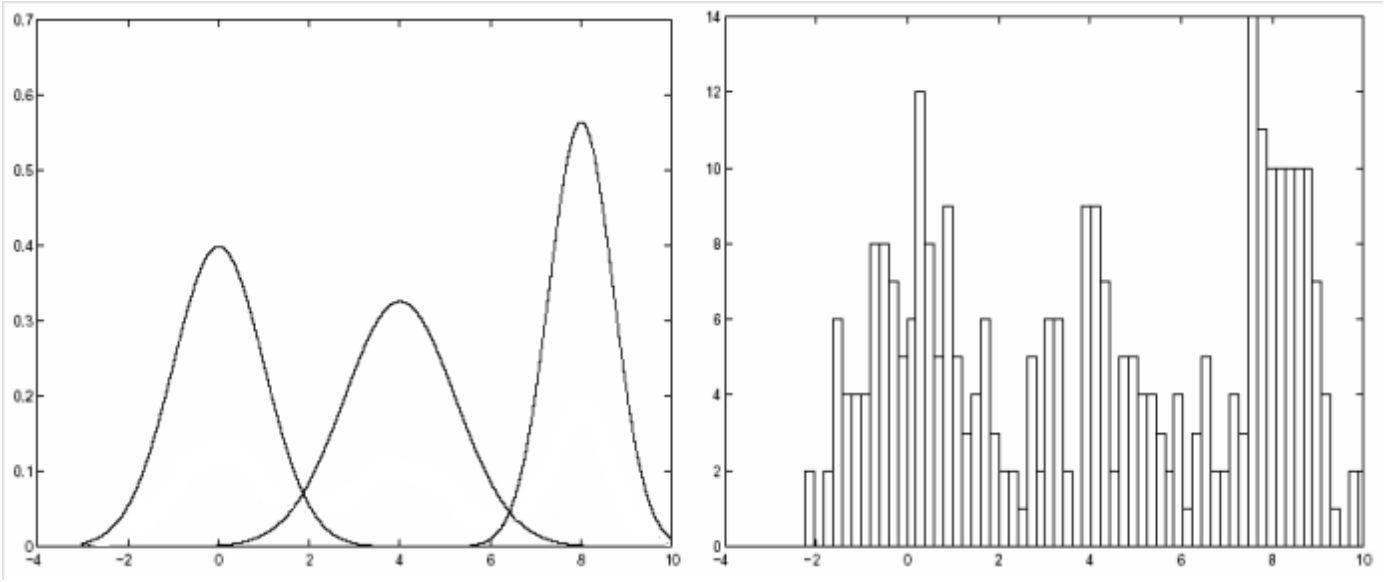


Figure. 2.7: (A gauche) un mélange de 3 gaussiennes présentées par leurs densités théoriques (A droite) l'histogramme associé [96].

2.3.2.3 Modèle basé sur une mixture de distributions beta

Il s'agit d'un modèle qui se base sur la distribution beta qui appartient à la famille de distribution des probabilités définie sur l'intervalle [0 1] avec deux paramètres de type réel positifs. Il a été démontré dans [25] et [144] que les modèles basées sur les mixtures de distributions beta (BMM) peuvent modéliser les données mieux que les modèles de mixture de gaussiennes. Un classifieur dédié à la détection des pixels peau et non peau basé sur ce type d'approche a été proposé par Zhanyu Ma et Arne Leijon dans [145]. Les auteurs ont modélisé la distribution des pixels peau dans le cadre de ce même modèle en exploitant l'espace de couleur RGB.
La fonction de la densité de probabilité de la distribution beta est définie comme suit:

$$Beta(x;u,v) = \frac{1}{beta(u,v)} x^{u-1}(1-x)^{v-1} \quad (2.36)$$

avec beta(u,v) est la fonction beta définie comme suit:

$$beta(u,v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)} \quad (2.37)$$

Où : $\Gamma(.)$ représente la fonction gamma défini comme suit: $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$.

Les données multivariées sont en général statistiquement dépendantes. Pour chaque vecteur x consistant en L éléments, les dépendances entre les éléments $x_1,\ldots x_L$ peuvent être représentées par un modèle de mixture.

L'auteur définit le modèle de mixture de distributions beta pour les données multivariées comme suit:

$$f(x;\Pi, U, V) = \sum_{i=1}^{I} \pi_i \text{Beta}(x; u_i, v_i) \quad (2.38)$$

où: $\text{Beta}(x; u_i, v_i) = \prod_{l=1}^{L} \text{Beta}(x_l; u_{li}, v_{li})$,

$\Pi = \{\pi_1,\ldots, \pi_I\}$: les poids associés à U et V.

$U = \{u_1,\ldots, u_I\}$ et $V = \{v_1,\ldots, v_I\}$

$\{u_i, v_i\}$ dénote les vecteurs de paramètres de la i^{ème} composante de mixture.

u_{li}, v_{li}: représentent les paramètres (scalaires) de la distribution beta pour l'élément x_l.

Afin de représenter la distribution de la couleur de la peau, chaque observation x est sous forme de vecteur tridimensionnel (L=3) et chaque élément x_l est dans l'intervalle [0, 1].

Afin d'estimer la distribution des paramètres du modèle de la mixture de distributions beta, les auteurs dans [145], ont défini d'abord la distribution à priori de la distribution beta comme suit:

$$f(u,v) = \frac{1}{C}\left[\frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)}\right]^{v_0} e^{-\alpha_0(u-1)} e^{-\beta_0(v-1)} \quad (2.39)$$

où: v_0, α_0, β_0 sont des paramètres positifs libres et C est un facteur de normalisation (en fonction de v_0, α_0, β_0) tel que:

$$\int_0^\infty \int_0^\infty f(u,v)\, du\, dv = 1 \tag{2.40}$$

On obtient la distribution postérieur de u, v comme suit :

$$f(u,v \mid X) = \frac{f(X \mid u,v) f(u,v)}{\int_0^\infty \int_0^\infty f(X \mid u,v) f(u,v)\, du\, dv} = \frac{\left[\frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)}\right]^{v_N} e^{-\alpha_N(u-1)} e^{-\beta_N(v-1)}}{\int_0^\infty \int_0^\infty \left[\frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)}\right]^{v_N} e^{-\alpha_N(u-1)} e^{-\beta_N(v-1)}\, du\, dv} \tag{2.41}$$

avec:

N : scalaire représentant le nombre d'observations X= {x_1,…..,x_N}.

$$\alpha_N = \alpha_0 - \sum_{n=1}^{N} \ln x_n$$

$$\beta_N = \beta_0 - \sum_{n=1}^{N} \ln(1 - x_n)$$

$$v_N = v_0 + N$$

L'estimation des paramètres de la distribution beta se fait de manière à avoir *f(u,v |X)* possède la même forme que *f(u,v)*. Dans le cas de l'équation (2.41), cette condition exige que le d'énumérateur soit égale à 1, autrement dit, l'équation suivante doit être satisfaite:

$$\int_0^\infty \int_0^\infty \left[\frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)}\right]^{v_N} e^{-\alpha_N(u-1)} e^{-\beta_N(v-1)}\, du\, dv = 1 \tag{2.42}$$

Le classifieur bayésien peau/non peau proposé dans [38] est appliqué par [145]. Les auteurs dans ce dernier ont entrainé 2 BMMs avec les pixels de couleur peau et non peau, chacun pour chaque type de peau. Etant donné un nouveau pixel x, la loi dénotant sa classification est la suivante:

$$\begin{cases} Si \dfrac{f(x \mid s)}{f(x \mid \sim s)} > \lambda . \dfrac{f(\sim s)}{f(s)} & \text{x} \in \text{s} \\ Sinon\, si \dfrac{f(x \mid s)}{f(x \mid \sim s)} < \lambda . \dfrac{f(\sim s)}{f(s)} & \text{x} \in \sim \text{s} \\ Sinon & \text{décision arbitraire} \end{cases} \tag{2.43}$$

avec: s et ~s représentent les classes peau et non peau respectivement. f(s) et f(~s) sont les probabilités à priori des couleurs peau et non peau. λ est une constante représentant un seuil à définir préalablement.

$$f(x\ |c) = \sum_{i=1}^{I} \pi_i \prod_{i=1}^{3} Beta(x_l; u_{li}^c, v_{li}^c), c \in \{s, \sim s\} \tag{2.44}$$

Nizar Bouguila et Djemel Ziou [23] ont proposé également un modèle de détection de la peau basé sur une mixture de distributions beta multidimensionnelle. Il s'agit de la Distribution de Dirichlet. On dit qu'un vecteur $\vec{X}(x_1, x_2 .. x_m)$ suit une distribution de Dirichlet si la fonction de densité de ce vecteur est donnée par l'équation suivante:

$$P(x_1, x_2, .. x_m) = \frac{\Gamma(\sum_{i=1}^{m+1} \alpha_i)}{\prod_{i=1}^{m+1} \Gamma(\alpha_i)} \prod_{i=1}^{m+1} x_i^{\alpha_i - 1} \tag{2.45}$$

avec α_i >0, ∀ i=1…m+1, $\sum_{i=1}^{m} x_i < 1$, $x_{m+1} = 1 - \sum_{i=1}^{m} x_i$ et 0< x_i<1, ∀ i=1,…,m.

La mixture Dirichlet avec m composantes est définie comme suit:

$$P(\vec{X} / \Theta) = \sum_{j=1}^{m} P(\vec{X} / j, \Theta_j) P(j) \tag{2.46}$$

Avec P(j) représente les proportions multidimensionnelles du modèle et Θ fait référence à l'ensemble des paramètres à estimer définis comme suit:
$\Theta = (\alpha_1, \alpha_{2,} .... \alpha_m, P(1), ..., P(m))$.

$\sum_{j=1}^{m} P(j) = 1$ et P(j) >0 ∀ j=1…m et Θ_j est représentée par le couple (α_j, P(j)).

L'estimation des paramètres se fait par maximisation de $P(\vec{X} / \Theta)$.

2.3.2.4 Modèle elliptique de borne

Après examen de la distribution de la couleur de la peau et de non peau dans plusieurs espaces de couleur, Lie et Yoo [136] ont constaté que cette distribution prend approximativement la forme d'une ellipse. Ils se sont basés sur une borne elliptique afin de séparer les pixels de peau et ceux de non peau dans le but de caractériser la véritable forme de la distribution des données d'apprentissage. Le modèle est défini comme suit:

$$\Phi(c) = (c - \phi)^T \sum\nolimits_{peau}^{-1} (c - \phi) \quad (2.47)$$

ϕ et Σ_{peau} sont calculés à partir des données d'apprentissage comme suit:

$$\phi = \frac{1}{|C_{peau}|} \sum_{c \in C_{peau}} c \quad (2.48)$$

$$\sum_{peau} = \frac{1}{N_{peau}} \sum_{c \in C_{peau}} N_{peau}(c)(c - \mu_{peau})(c - \mu_{peau})^T \quad (2.49)$$

où $| C_{peau} |$ est le nombre de pixels de couleur de peau avec $C_{peau} \subseteq C$.
μ_{peau} est l'espérance des vecteurs d'apprentissage de pixels de peau défini par l'équation (2.33). Pour classifier les pixels, on compare $\Phi(c)$ avec un seuil fixé θ. c est considéré comme un pixel de peau si $\Phi(c) < \theta$ sinon il est un pixel de non peau. Les tests de classification effectués par ce modèle sur six espaces de couleur testés sont satisfaisants. De plus, cette méthode est plus rapide lors de la classification que les méthodes basées sur une simple gaussienne [136].

2.3.3 Modèles de peau non paramétriques

Les approches non paramétriques ne dépendent pas de la forme de la fonction de distribution de la teinte.
Les approches non-paramétriques visent à estimer la distribution de la couleur de la peau sans une modélisation explicite à partir d'un ensemble d'apprentissage.
Ce type d'approche revient à estimer la distribution de la couleur de peau des données d'apprentissage sans aucune hypothèse sur la densité de distribution colorimétrique des pixels traités.

2.3.3.1 Classification de pixels par table de correspondance: *lookup table (LUT)*

Cette méthode se base sur le calcul d'histogramme à partir d'exemples afin de construire un modèle de peau. Son principe consiste à associer à chaque pixel d'une image sa probabilité d'appartenance à une instance du modèle constituant ainsi une carte de correspondance. A partir de cette carte, les pixels sont classés. Plusieurs algorithmes de détection et de suivi de visage se sont appuyés sur cette méthode pour segmenter les pixels de peau ([226], [40], [88], [182], [185], [189] et [20]). Après projection dans un espace de couleur bien choisi, le modèle prend la forme d'un histogramme des valeurs des pixels de la base d'exemples. La quantification de l'espace de couleur se fait par un nombre de cases (bins) formant un histogramme de couleur 3D ou 2D. A chaque case des axes lui correspond le nombre d'occurrences que la valeur de couleur s'est produite dans les images de peau de la base d'apprentissage. Ces valeurs de couleurs forment alors une table de correspondance dite "Lookup Table" avec les occurrences de leur apparition.

Les histogrammes (2D) demeurent les plus utilisés dans la littérature. Nous citons comme exemple l'espace (r, g) (composantes R et G normalisées) par Bérard [15], l'espace (H, S) (teinte et saturation) par Wu [219]. A partir d'un histogramme tridimensionnel (3D), on peut déduire, par projection de l'histogramme (3D) sur 2 des 3 plans colorimétriques d'une image I, trois histogrammes (2D). Une fois l'histogramme construit, il est normalisé et ces valeurs sont converties en distribution discrète P(c/peau).

$$P(c / peau) = \frac{peau(c)}{Norm} \tag{2.50}$$

où *peau(c)* est le nombre de pixels associé à une case (bins) de l'histogramme de peau formé par le vecteur de couleur c et *Norm* est le coefficient de normalisation.
Ce coefficient a été défini de différentes manières. Il correspond au nombre total de pixels de peau dans les travaux de [116] et [169]. Il s'agit de la plus grande valeur dans la table de correspondance (Lookup Table) dans les travaux de [226]. Les valeurs normalisées de cette table constituent la probabilité qu'une couleur corresponde à la peau. Un pixel est considéré comme pixel de peau si sa valeur dépasse un certain seuil.

L'histogramme de chrominance est un modèle fiable pour la reconnaissance d'entités colorées selon Swain et Ballard [193]. Ils expérimentent différents types d'histogrammes dont ceux créés à partir des composantes rouges et vertes normalisées. Chaque cellule de coordonnées (r, g) de l'histogramme à deux dimensions h_E donne le nombre de pixels de l'échantillon E ayant une chrominance de composante rouge r et verte g. Cet histogramme permet de définir la probabilité de l'équation (2.50) par:

$$P(c / peau) = \frac{1}{n_E} h(c) = \frac{1}{n_E} h(c_r, c_g) \tag{2.51}$$

où n_E est le nombre total de pixels de l'ensemble E.
Afin de classifier les pixels en pixels de peau ou non peau, Cai et al. [33] se basent sur une table de correspondance construite selon les deux axes de couleur a et b de l'espace CIE L*a*b*. L'établissement de la table a été faite à partir de 2300 échantillons de peau extraits de 80 images contenant des pixels de peau [96].

2.3.3.2 Modèle basé sur l'appariement d'histogrammes

L'histogramme de couleur illustre la fréquence d'apparition de chaque couleur. Ce type d'information est global et néglige la disposition spatiale des pixels dans l'image. Saxe et Foulds [179] exploitent l'espace de couleur HSV et une technique basée sur l'appariement d'histogrammes, décrite dans [193]. Les auteurs proposant une sélection d'une région de peau, dite région de contrôle, qui est comparée par la suite au reste de l'image en utilisant l'appariement d'histogramme. Cela revient à comparer l'histogramme de la région de contrôle avec les histogrammes des régions de même

taille de l'image. D'abord, l'algorithme transforme l'image d'entrée dans l'espace de couleur HSV, puis l'utilisateur choisit manuellement la région de contrôle (germe initial). Ensuite, l'image est examinée région par région en utilisant la méthode d'intersection d'histogramme (équation 2.52).

$$M_{C,I} = \frac{\sum_{i,j} \min(H^C(i,j), H^I(i,j))}{\sum_{i,j} H^C(i,j)} \qquad (2.52)$$

Les histogrammes sont calculés en se basant sur les valeurs de teinte et de saturation. Un bloc est considéré comme peau si le score d'appariement ($M_{C,I}$) entre l'histogramme de contrôle (H^C) et l'histogramme correspondant à un bloc de l'image (H^I), excède un seuil donné. Une fois que tous les blocs de l'image ont été examinés, l'identification de blocs de peau sera améliorée par élimination de blocs annexes. On entend par "bloc annexe" une région dont aucun de ses huit voisins n'a été marqué comme peau. Finalement, l'algorithme choisit un nouveau bloc de contrôle parmi les blocs de peau qui ont été conservés (ne comprenant pas les blocs retirés). Le processus sera répété jusqu'à ce qu'aucun nouveau bloc n'ait été identifié pendant l'itération en cours. Dans ce cas, le processus d'identification des blocs de peau est considéré comme achevé. Les auteurs ne fournissent pas beaucoup de détails sur la façon de choisir un nouveau bloc de contrôle. Deux méthodes possibles sont envisageables. La première consiste à choisir comme nouveau bloc celui qui correspond au score d'appariement le plus élevé ($M_{C,I}$), alors que pour la deuxième méthode, le choix du nouveau bloc de contrôle se porte sur celui qui correspond au score le moins élevé.

Cependant, plusieurs problèmes ont été rapportés sur cet algorithme. D'abord, le choix du bloc initial de contrôle est parfois trop critique. Parfois, en utilisant un même seuil, on peut avoir deux résultats différents si l'emplacement de la sélection initiale diffère seulement par deux pixels. En considérant deux sélections de blocs de contrôle qui ne différent que de deux pixels, on constate qu'on peut avoir des résultats très différents. Ainsi, cet algorithme exige une bonne sélection initiale du bloc de contrôle. Ensuite, le choix d'un seuil adéquat est parfois difficile. En effet, si le choix d'un premier seuil X permet de localiser un nombre limité de blocs considérés comme blocs de peau, un deuxième seuil formé par une légère modification du premier tel que X_2= X-0.01 permet d'identifier un nombre important de blocs de peau. On constate alors qu'un insignifiant changement du seuil conduit à un grand changement des résultats. La cause de ces deux problèmes n'a jamais été déterminée.
Cet algorithme a donné suite a des résultats concluants dans certains cas, alors qu'il a mal fonctionné dans d'autres cas [96].

2.3.3.3 Modèle Bayésien basé sur les Histogrammes

Le modèle de couleur de peau et de non peau représenté dans [117] et [37] se base sur des histogrammes. L'espace de couleur C est quantifié à un certain nombre

de bins c ϵ C et le nombre de pixels de couleur dans chaque bin est compté. Ils normalisent chaque bin pour obtenir la distribution conditionnelle des pixels de peau et de non peau P(c/peau) et
P(c/non peau).
Supposons la notation suivante dans la base d'apprentissage:

- $N_{peau}(c)$: le nombre de pixels de couleur c pour la classe de peau.
- $N_{\neg peau}(c)$: le nombre de pixels de couleur c pour la classe de non peau.
- N_{peau}: le nombre total de pixel peau.
- $N_{\neg peau}$: le nombre total de pixels non-peau.

Nous avons donc:

$$P(c / peau) = \frac{N_{peau}(c)}{N_{peau}} \tag{2.53}$$

$$P(c / \neg peau) = \frac{N_{\neg peau}(c)}{N_{\neg peau}} \tag{2.54}$$

$$P(peau) = \frac{N_{peau}}{N_{peau} + N_{\neg peau}} \tag{2.55}$$

$$P(\neg peau) = \frac{N_{\neg peau}}{N_{peau} + N_{\neg peau}} = 1 - P(peau) \tag{2.56}$$

Ensuite, la formule de Bayes ci dessous est utilisée pour calculer la probabilité qu'un pixel soit un pixel peau ou non selon sa couleur.

$$P(peau / c) = \frac{P(c / peau)P(peau)}{P(c / peau)P(peau) + P(c / \neg peau)P(\neg peau)} \tag{2.57}$$

$$P(\neg peau / c) = 1 - P(peau / c) \tag{2.58}$$

Un seuil θ est alors établit afin de délimiter les deux classes tel que $0 < \theta < 1$. Le pixel est reconnu comme pixel peau si P(peau/c) > θ et comme pixel non peau si P(peau/c) <=θ [96].

2.3.3.4 Les réseaux de neurones

Les réseaux de neurones sont utilisés pour classifier les pixels d'une image en classes peau et non peau. Un réseau de neurones multi couche MLP (multi layer

perception) est proposé par Ming Jung et al. en 2003 dans [152]. Ils ont utilisé 41 000 pixels de peau dans l'espace RGB soumis à différentes conditions d'illumination pour l'apprentissage afin d'identifier les différentes parties du visage. Karlekar et Desai dans [122] ainsi que Phung et al. dans [164] ont utilisé un MLP dans l'espace CbCr pour la classification de la peau. Le MLP est entrainé à partir de 200 images en ayant recours à l'algorithme *Levenberg-Marquardt* pour une convergence plus rapide. Ming-Jung et al. dans [152] ainsi que Zhu et al. dans [230] ont entrainé des réseaux de neurones à trois couches dans l'espace RGB pour extraire les régions peau. Il existe dans la littérature deux types de modèles basé sur la détection de la peau [27]: les modèles *symétriques* et les modèles *asymétriques*. Les modèles symétrique utilisent un seul classifieur tandis que les modèles asymétrique utilisent deux classifieurs distincts l'un dédié pour les pixels peau, l'autre pour les pixels non peau. L'avantage avec les modèles asymétriques est qu'ils améliorent la séparation entre les pixels peau et ceux de non peau avec l'inconvenient de la complexité élevé du temps nécessaire pour entrainer deux classifieurs distincts. Les auteurs dans [17] ont utilisé une approche symétrique en utilisant un réseau de neurones MLP dans l'espace RGB avec deux neurones dans la couche de sortie: l'un dédié à la classe peau et l'autre à la classe non peau. L'approche adoptée dans [17] permet d'améliorer la séparation entre les classes peau et non peau aussi bien que le font les approches basées sur les modèles asymétriques en apportant un net avantage par rapport à ces derniers qui est la réduction considérable du temps nécessaire pour l'apprentissage. La structure du réseau proposé est présentée sur la figure 2.8. Le réseau est constitué de trois couches. Une couche d'entrée est composée de trois nœuds, chacun correspondant à une composante de l'espace RGB. Une couche intermédiaire comportant 5 nœuds et une couche de sortie comportant deux nœuds, l'un dédié pour la classe peau et l'autre pour la classe non peau. Un seuil pour la séparation des deux classe est introduit noté θ tel que $0< \theta\leq 2$. La classification est établit comme suit:

Pour un nouveau pixel à classifier,

Si $C_1- C_2>=0$ alors le pixel est un pixel peau

Sinon il s'agit d'un pixel non peau.

L'idéal est d'avoir $C_1-C_2 = 2$ pour la classe peau et $C_1-C_2 = -2$ pour la classe non peau. Le classifieur ainsi construit est invariant aux changements d'illumination et applicable pour les différents types de couleur de peau.

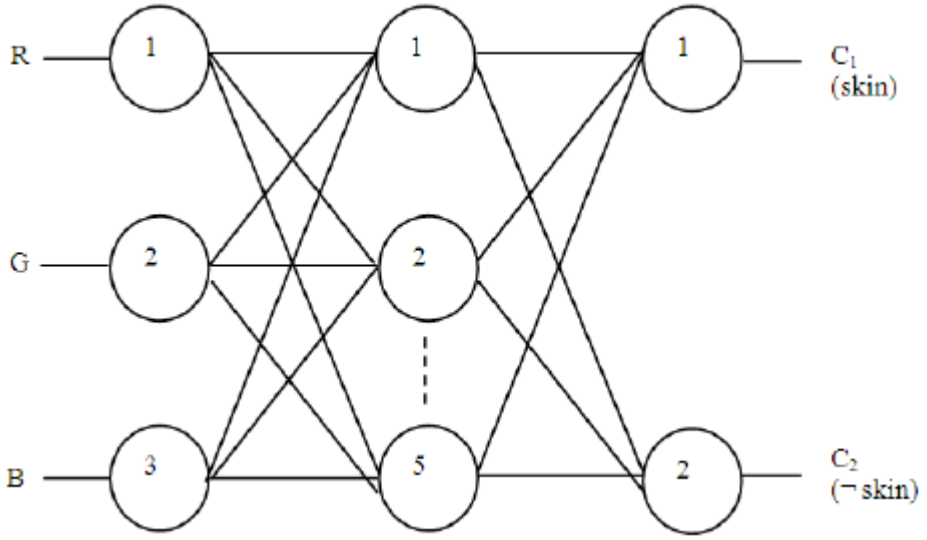


Figure. 2.8: Architecture du réseau de neurones exploité dans [17].

2.3.3.5 Self organizing map (*SOM*)

L'algorithme *SOM* [126] ou algorithme de Kohonen, est l'un des plus célèbres et des plus utilisés des réseaux neuronaux artificiels non supervisés [194]. Brown et al. [27] se sont basés sur l'algorithme (SOM) pour identifier les pixels de peau dans l'image dans leur système de détection de visages. Ils ont établis deux SOMs, le premier ne représente que des pixels de peau, appris à partir de 30000 pixels de peau, alors que le deuxième représente des pixels de peau et de non peau où 15000 pixels de chaque classe ont été utilisés.

Le principe de la carte auto adaptative de Kohonen est comme suit:

Soit: $\Gamma=\{X_1, X_2,\ldots\ldots, X_q,\ldots.., X_Q\}$ un échantillon constitué de Q observations définies dans un espace à N dimensions telles que: $X_q = [X_{q,1},X_{q,2},\ldots,X_{q,n}\ldots\ldots,X_{q,N}]^T$, q=1, 2,…, Q.
La structure du réseau de Kohonen est constituée de deux couches:

- La couche d'entrée: composée de N neurones représentants les N attributs d'une observation X_q.

- La couche de sortie: composée d'un nombre M de neurones régulièrement répartis sur une carte.

Les neurones de la couche d'entrée sont connectés à ceux de la couche de sortie. Chaque connexion d'un neurone d'entrée j vers un neurone de sortie m a le vecteur poids $W_{m,j}$. Ainsi chaque neurone de sortie m a le vecteur poids:

$W_m=[W_{m,1}\ W_{m,2}\ldots.W_{m,n}\ldots.W_{m,N}]^T$

Chaque neurone de la couche de sortie est défini par sa position relative sur la carte et par son vecteur poids dans l'espace de représentation des observations.

L'apprentissage du réseau vise, à chaque présentation à l'itération t d'une observation $X_q(t)$ à l'entrée du réseau, à déterminer le neurone gagnant, c'est à dire celui dont le vecteur poids est le plus proche de cette observation. Le vecteur poids du neurone gagnant et ceux des neurones voisins sur la carte subissent alors des modifications en fonction de l'observation présentée au réseau.

Cette structuration des neurones et cette technique d'apprentissage permettent aux neurones voisins sur la carte d'être sensibles à des observations voisines dans l'espace d'origine: c'est le phénomène d'auto-adaptation décrit par Kohonen [127]. A la fin de cette étape, chaque neurone devient sensible à une zone de l'espace de représentation des observations et son vecteur poids converge vers le barycentre des observations présentes dans cette zone.
Le voisinage de rayon r d'un neurone m est défini par l'ensemble des neurones m' tel que:

$$V(m,r) = \{m' \in [0, M[, m' \neq m / d(U_m, U_{m'}) \leq r)\} \tag{2.59}$$

où $d(U_m,U_{m'})$ désigne la distance Euclidienne entre les vecteurs U_m et $U_{m'}$ qui sont les positions respectives des neurones m et m' sur la carte.
A la présentation au réseau de chaque observation $X_q(t)$, le vecteur poids du neurone gagnant, noté m^*, et ses voisins sont modifiés pour chaque itération t tels que:

$$\begin{cases} W_m(t) = W_m(t-1) + a(t).[X_q(t) - W_m(t-1)] & si\ m = m^* \\ W_m(t) = W_m(t-1) + a(t).h(m^*,t)[X_q(t) - W_m(t-1)] & si\ m \in V(m^*, r(t)) \\ W_m(t) = W_m(t-1) & si\ m \notin V(m^*, r(t))\ \ et\ \ m \neq m^* \end{cases} \tag{2.60}$$

où:

- $W_m(t)$: est le vecteur poids du groupe de connexions des neurones de la couche d'entrée vers le $m^{ème}$ neurone de la couche de sortie, à l'itération t.
- r(t) est le rayon de voisinage ou d'interaction qui dépend du rang t de l'itération considérée.
- a(t) est le coefficient d'apprentissage. Il peut être une fonction hyperbolique, une fonction exponentielle ou aussi une fonction linéaire du paramètre d'itération t.
- m^* est le neurone gagnant défini par:

$$m^* = Arg \min_m [d(X_q(t), W_m(t))] \tag{2.61}$$

- $h(m^*,t)$ est la fonction d'interaction défini par:

$$h(m^*, t) = \exp\left(- \frac{d^2(U_m, U_{m^*})}{2r^2(t)} \right) \tag{2.62}$$

Brown et al. [27] ont opté pour un espace de couleur à deux dimensions (r, g). L'architecture de la carte pourrait être rectangulaire ou hexagonale, mais ils ont choisi de travailler sur une architecture hexagonale. La figure 2.9 [96] illustre un exemple de deux cartes ainsi que les voisins de leurs nœuds centraux. Dans le but de comparer ce modèle avec d'autres modèles, les auteurs ont établis une série de tests et ont prouvé que l'algorithme SOM est légèrement meilleur que le modèle de mélange de gaussienne, alors qu'il est inférieur au modèle basé sur les histogrammes au niveau de la classification de pixels. L'avantage de cette méthode est qu'elle consomme moins de ressource que les 2 méthodes citées [96].

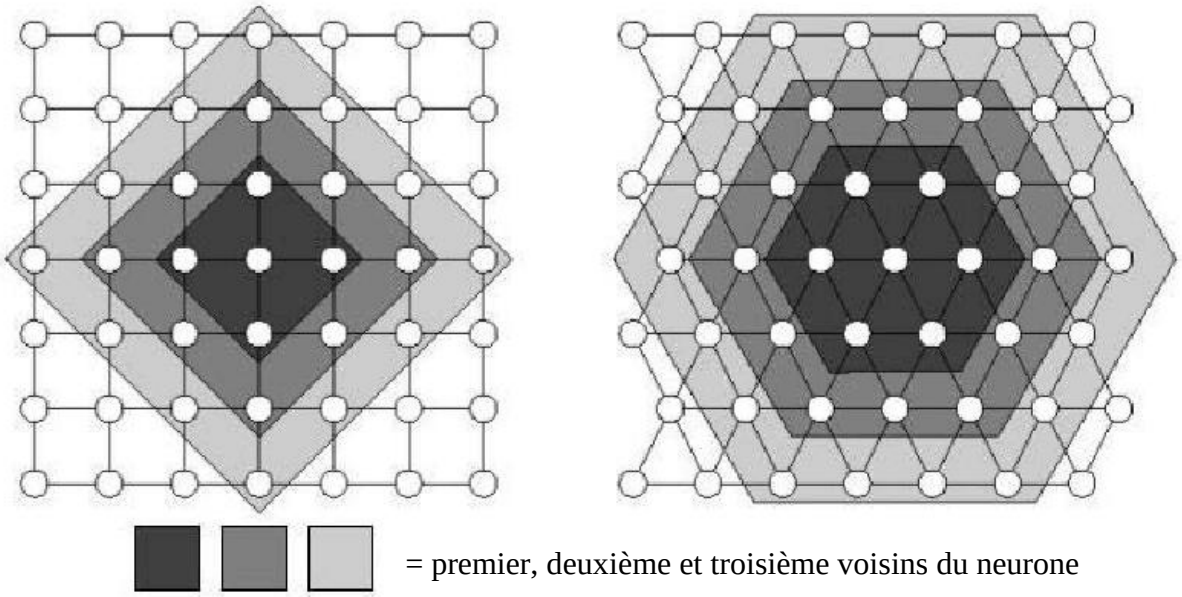


Figure. 2.9: Architectures possibles de la carte SOM: rectangulaire (à gauche) et hexagonale (à droite).

2.3.3.6 Combinaison de réseaux de neurones pour la détection de la peau

Le seul travail à avoir combiné différents classifieurs de peau et non peau basés sur les réseaux de neurone a été effectué par Chelsia Amy Doukim et al. dans [66] en exploitant l'espace YCbCr. Ces derniers ont eu recours à plusieurs stratégies afin de combiner des réseaux de neurones, chacun dédié à la classification des pixels peau et non peau.
Plusieurs composantes de chrominance extraites de l'espace YCbCr ont été utilisées dans le cadre de ce travail comme caractéristiques colorimétriques de la peau.

L'une des caractéristiques importantes d'un réseau de neurones est la définition de sa structure. Le nombre de perceptrons de la couche d'entrée est lié au nombre de composantes de chrominance utilisées. Les auteurs ont utilisés les composantes suivantes: Cb, Cr, Cb/Cr, Cb.Cr et Cb-Cr. Les perceptrons de la couche de sortie

servent de critères de décision et leur nombre est égale au nombre de classes de sortie qui est une seule dans ce cas: la classe peau. Dans [66], on a utilisé une seule couche intermédiaire en se référant à ce qui a été cité dans [80] où les auteurs ont montré qu'une seule couche intermédiaire est suffisante pour résoudre plusieurs problèmes pratiques. Les auteurs ont utilisé l'annotation *C-HN-O* pour désigner une topologie du réseau de neurones à trois couches. *C* représente le neurone d'entrée et correspond à la composante de chrominance à utiliser. *HN* (*hidden layer*) représente la couche intermédiaire et correspond au nombre de neurone de cette couche. *O* (*output*) correspond au neurone de sortie. Ce neurone est étiqueté par une variable binaire (1 pour indiquer la classe peau et 0 pour indiquer la classe non peau).

Afin de déterminer le nombre de neurones de la couche intermédiaire, une méthode appelée "*coarse to fine search*" est appliquée selon laquelle les valeurs de HN à tester sont 1, 2, 4, 8, 16, 32, 64 et 128. Chaque réseau avec une HN donnée est entrainé 30 fois en utilisant différentes valeurs d'entrée et l'erreur moyenne à travers les 30 entrainements est calculée. Ainsi la valeur HN qui donne l'erreur la plus petite est sélectionnée. Par la suite, afin d'obtenir le nombre optimal du nombre de neurones de HN, un intervalle de recherche séquentielle est défini. Si la valeur HN choisie durant la phase précédente est HN_B, alors parmi les valeurs qui lui sont inférieures, on note celle qui lui est la plus proche par HN_{BL} et parmi les valeurs qui lui sont supérieurs, on note celle qui lui est la plus proche par HN_{BH}. La définition de l'intervalle de la recherche séquentielle est défini comme suit:

Soit N : le nombre de nœuds de HN.
I : intervalle de recherche séquentielle.

Si $HN_B = 2$

$I = [HN_{BL}, \quad HN_B + (1/2 * (HN_B - HN_{BL}))]$

Sinon

Si HNB =128

$I = [HN_{BL} + (1/2 * (HN_B - HN_{BL})), \quad HN_B]$

Sinon

$I = [HN_{BL} + 1/2 * (HN_B - HN_{BL}), \quad HN_B + (1/2) * (HN_{BH} - HN_B)]$

Le but est de trouver la valeur la plus faible qui connait la valeur de l'erreur la plus faible. La figure suivante illustre un exemple où les valeurs HN_B , HN_{BL} et HN_{BH} sont respectivement 64, 32 et 128.

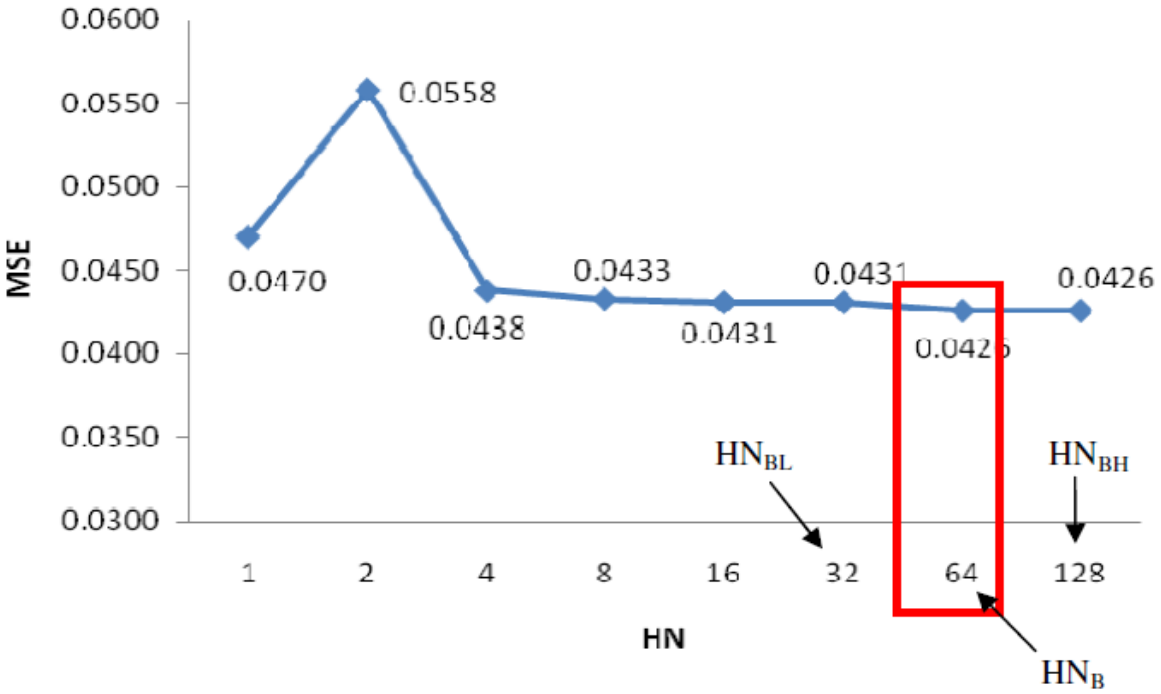


Figure. 2.10: Sélection de la valeur HN à laquelle correspond la valeur de l'erreur la plus basse [66].

L'intervalle de recherche séquentielle est donc [48 96]. La figure suivante montre le résultat de recherche séquentielle dans cet intervalle.

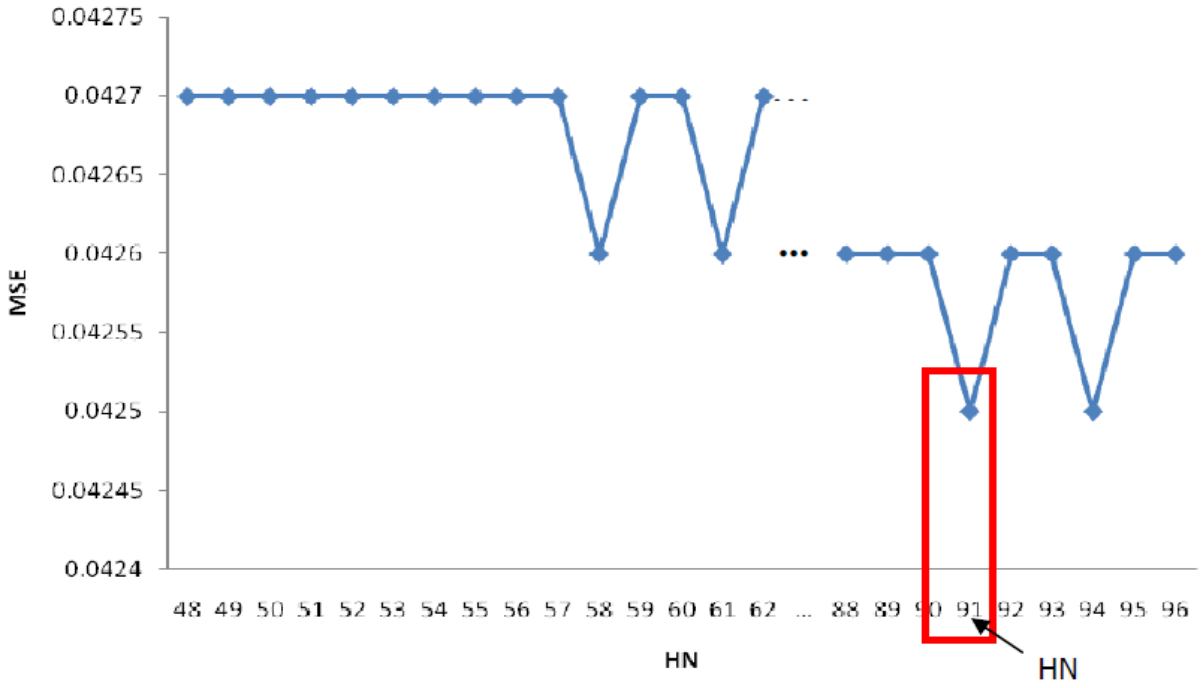


Figure. 2.11: Recherche séquentielle dans l'intervalle [48 96] [66].

La valeur retournée par la recherche est 91. Le réseau de neurones aura donc la forme suivante: 1-91-1.
Parmi les types de combinaisons de réseaux de neurones recensés dans [66] :

a. Combinaison des sorties des classifieurs peau en utilisant les opérateurs logique 'and' et 'or': Selon la figure 2.12 les combinaisons possibles des deux classifieurs présentés sont 'Cb-Cr and Cb/Cr' et 'Cb-Cr or Cb/Cr'. Le seuil de classification peau/ non peau est représenté par la valeur 0.5.

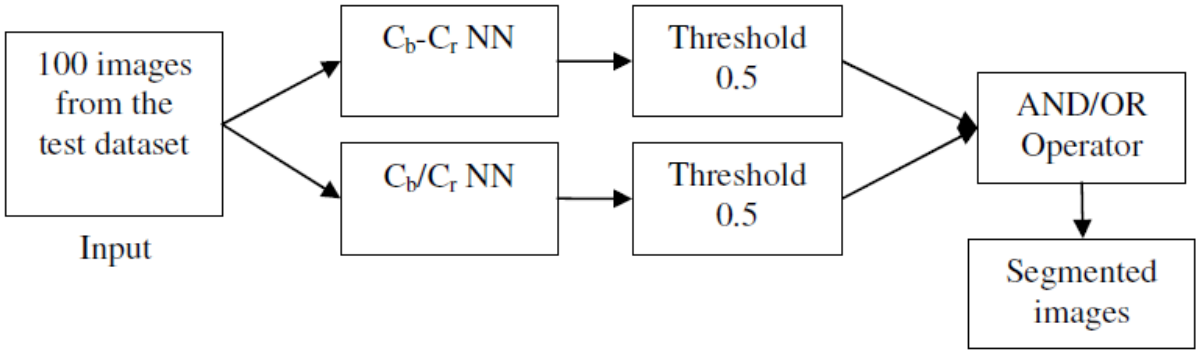


Figure. 2.12: Combinaison des sorties de classifieurs en utilisant les opérateurs binaire [66].

b. Combinaison des sorties des classifieurs peau en utilisant le vote: La décision de classification dépend du nombre de vote majoritaire des différents classifieurs. Si la majorité des classifieurs considèrent que le nouveau pixel à classer est un pixel peau alors il sera classifié comme tel, autrement il ne l'est pas.

c. Combinaison des sorties des classifieurs peau en utilisant la pondération des entrées et l'addition des sorties: Dans ce cas, des poids choisis expérimentalement sont affectés aux entrées des différents classifieurs de tel manière à avoir la somme des poids est égale à 1. La somme des sorties des classifieur est comparée à un seuil afin d'effectuer la classification. La figure suivante illustre un exemple.

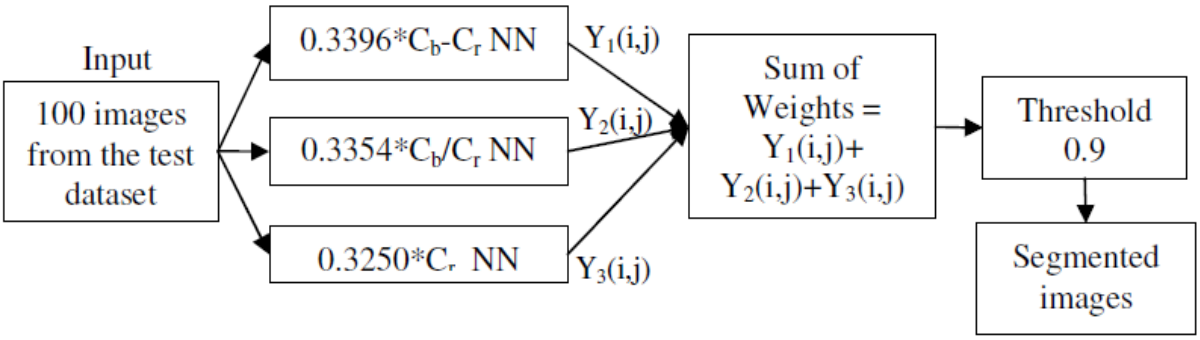


Figure. 2.13: Combinaison des sorties des classifieurs en utilisant la pondération des entrées et l'addition des sorties [66].

2.3.4 Modèles flous

Outre les modèles explicites, paramétriques et non paramétriques, on assiste à l'émergence de techniques populaires en détection de la peau connus sous le nom de modèles flous. Ces derniers se basent sur la logique floue pour décrire les entrées du système qui peuvent être dans notre cas les composantes d'un espace de couleur utilisé lors de la phase du prétraitement d'une image. A l'inverse de la logique booléenne, la logique floue permet à une condition d'être en un autre état que *vrai* ou *faux*. Il y a des degrés dans la vérification d'une condition. Plusieurs approches basées sur la logique floue ont été rapporté dans le cadre de la détection de la peau.

2.3.4.1 Système d'inférence flou Takagi Sugeno

Hamid et Jemma on proposé en 2006 un système d'inférence flou *Takagi Sugeno* pour la classification des pixels [98]. Ce système consiste en deux variables d'entrée qui sont les composantes de chrominance de l'espace de couleur YCbCr (Cb et Cr) et une seule variable de sortie qui indique l'appartenance soit à la classe peau ou non peau. Chaque variable d'entrée peut appartenir à l'une des catégories suivantes: clair, moyen et sombre. A partir de cette catégorisation, des règles Si-Alors sont appliqués pour chaque pixel de l'image afin de prendre une décision sur l'affectation du pixel soit à la classe peau désignée par la valeur 1 ou non peau désignée par la valeur 0. Les règles floues appliquées sont:

1. Si Cb est clair et Cr est clair, alors pixel = 0.
2. Si Cb est clair et Cr est moyen, alors pixel = 0.
3. Si Cb est clair et Cr est foncé, alors pixel = 0.
4. Si Cb est moyen et Cr est clair, alors pixel = 0.
5. Si Cb est moyen et Cr est moyen, alors pixel = 1.
6. Si Cb est moyen et Cr est foncé, alors pixel = 1.
7. Si Cb est foncé et Cr est clair, alors pixel = 0.
8. Si Cb est foncé et Cr est moyen, alors pixel = 1.
9. Si Cb est foncé et Cr est foncé, alors pixel = 0.

Le degré d'appartenance à des ensembles flous appropriés est déterminé par des fonctions spécifiques. La décision finale du système d'inférence correspond à la moyenne des décisions des règles présentées en dessous pondérées chacune par un poids P_i.

2.3.4.2 Algorithme C-Mean flou modifié (CMFM)

Une méthode de clustering flou modifié a été proposée par Chahir et Elmoataz en 2006 [39] pour classifier les pixels d'une image en C classes dans l'espace HSV en minimisant la fonction suivante:

$$J = \sum_{i=1}^{C}\sum_{j=1}^{N}(u_{ij})^2 d^2(x_j, c_i) - \alpha\sum_{i=1}^{C} p_i \log(P_i)\,. \qquad (2.63)$$

avec:

$$P_i = \frac{1}{N}\sum_{j=1}^{N} u_{ij} \tag{2.64}$$

N est le cardinal de tous les clusters réunis. P_i est interprété comme la probabilité du cluster i et α un paramètre à fixer. $d_{ij}^2 = d^2(x_j, c_i)$ est la distance euclidienne standard entre le centroide du i^{ème} cluster (c_i) et du j^{ème} pixel.
Dans le but de tenir compte de la distribution de la teinte, la pureté, la luminosité et la distribution spectrale des pixels, la fonction distance est définie comme suit:
$d_{ij} = \sum_{k=1}^{4} w_k d(x_{jk}, c_{ik})$ et $P_i = \sum_{k=1}^{4} \frac{N_{ik}}{N}$. w_k sert de poids constant correspondant au canal k.

Le calcul du centroide de la classe j se fait comme suit:

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}.x_i}{\sum_{i=1}^{N} u_{ij}} \tag{2.65}$$

c_{ik} et x_{jk} représentent respectivement le centroide de la classe i dans le canal k et le j^{ème} pixel dans le même canal. N_{ik} correspond au cardinal du cluster i dans le canal k.
Le calcul du centroide de la classe j se fait comme suit:

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}.x_i}{\sum_{i=1}^{N} u_{ij}} \tag{2.66}$$

Les élément de la matrice $U=[u_{ij}]$ se calcule selon la formule suivante:

$$u_{ij} = \frac{\frac{1}{d_{ij}^2}}{\sum_{k}^{C}(\frac{1}{d_{kj}^2})} + \frac{\alpha}{2Nd_{ij}^2}\left[\log P_i - \frac{\sum_{k}^{C}(\frac{1}{d_{kj}^2 \log P_k})}{\sum_{k}^{C}\frac{1}{d_{kj}^2}}\right] \tag{2.67}$$

Le processus est réitéré lors de la minimisation de (2.63) jusqu'à avoir $|J^n - J^{n+1}| < \square$ où $\square$ est un seuil d'arrêt de boucle à définir préalablement. L'algorithme itératif

dénotant ce processus est présenté comme suit où k est une variable correspondant à la $k^{ième}$ itération.

1. Initialiser la matrice $U=[u_{ij}]$. Initialiser J.
2. Calculer les centroides des C classes considérées $C=[c_j]$.
3. mise à jour de la matrice U^k, U^{k+1} en tenant compte des centroide calculés dans l'étape 2.
4. Mise à jour de J^k, J^{k+1}.
5. si $|J^{k+1}- J^k|<$ □ alors stop, sinon aller à (2).

Une fois les centroides des classes calculés suite à l'algorithme précédent, un pixel est affecté à la classe dont le centroide est lui le plus proche par la distance euclidienne [39].

2.3.5 Les modèles hybrides

2.3.5.1 Clustering soustractif flou et règles explicites

Un modèle combinant un système basé sur la logique floue avec un autre basé sur la définition de règles explicites dans l'espace colorimétrique RGB a été proposé par A.A. Zaidan et al. en 2010 [225]. La génération de règles en logique floue se base sur un clustering de type soustractif développé par Chiu en 1994 [43]. Afin d'extraire les règles floues à partir des données, le clustering soustractif agit sur l'espace entrée/sortie de chaque classe pour identifier l'affectation de nouvelles données. Chaque pixel est représenté sous forme d'un vecteur contenant à la fois les variables d'entrée et de sortie. Afin d'établir une bonne classification de nouvelles données, les caractéristiques d'entrée/sortie spécifique à chaque cluster sont étudiées à partir de son centroide qui est utilisé comme un prototype représentant le cluster auquel il est associé. Considérons un ensemble de n pixel $\{x_1, x_2,...,x_n\}$.
Au départ chaque point de l'espace RGB x_i est considéré comme un possible centroide d'un cluster dont son potentiel est calculé comme suit:

$$P_i = \sum_{j=1}^{n} \exp(-(\frac{2}{r_a})^2 \left\| x_i - x_j \right\|^2) \qquad (2.68)$$

où r_a est une constante et ||.|| représente la distance euclidienne. Chiu [44] a proposé que la valeur de r_a soit entre 0.2 et 0.5, tandis que Hammouda et Karray [97] ont montré qu'elle est définie entre 0.4 et 0.7. Le pixel qui possède la plus grande valeur de son potentiel est désigné comme le premier centroide. Soient x_1^* le premier centroide et P_1^* son potentiel calculé. Le potentiel de chaque point de l'espace RGB est mise à jour comme l'indique la formule suivante:

$$P_i \Leftarrow P_i - P_1^* \exp(-(\frac{2}{r_b})^2 \left\| x_i - x_1^* \right\|^2) \qquad (2.69)$$

où r_b est une constante. Un bon choix de r_b est r_b= 1.5 r_a [43] ou r_b=1.25 r_a [44]. Le processus est réitéré jusqu'à avoir le potentiel de tout les autres points ne faisant pas parti de l'ensemble des centroides soit inférieur à un seuil calculé en fonction de P_1^*.

Soit un ensemble de m centroides chacun représentant un cluster $\{x_1^*, x_2^*, ..., x_m^*\}$. Chaque centroide est représenté sur un espace de M dimensions dont N correspondent aux variables d'entrée et M-N aux variables de sortie. Chaque vecteur x_i^* est constitué de deux composantes y_i^* et z_i^* où y_i^* contient les N premières composantes de x_i^* et z_i^* contient les M-N restantes. Afin d'identifier le comportement du système en termes de classification, on doit extraire dans chaque centroide une règle floue qui est sous la forme suivante :

Règle i:

Si {entrée est proche de y_i^*} alors {sortie est proche de z_i^*}.

Etant donné un vecteur y en entrée, le degré de proximité de y à y_j^* est défini comme suit:

$$\mu_j = \exp(-(\frac{2}{r_a})^2 \left\| y - y_j^* \right\|^2)$$

(2.70)

Par la suite, les pixels affectés à la classe peau son soumis à un nouveau test basé sur des règles explicites appliqués sur les composantes de l'espace de couleur de présentation (RGB). Si les conditions imposées par toutes les règles sur un nouveau pixel à tester sont vérifiées alors il s'agit d'un pixel peau sinon il ne l'est pas. Ces règles sont supposées selon les auteurs dans [225] améliorer les performances du classifieur de peau et non peau. Un pixel est considéré comme peau si toutes les règles suivantes sont vérifiées:

Règle 1: L'intervalle de définition des valeurs des composantes de l'espace RGB ne doit pas être étroit.
R>95 et G>40 et B>20 et max{R,G,B}-min{R,G,B}>15.

Règle 2: Les composantes R et G ne doivent pas être proches l'une de l'autre.

|R-G|>15.

Règle 3: La valeur de la composante R doit être plus grande que celle de G et B car la peau est constituée de sang et de mélanine si la peau n'est pas soumise à l'éclairage du flash de l'appareil d'acquisition de l'image ou bien à l'illumination latérale du jour.

R>G et R>B.

Règle 4: R>220 et G>210 et B>170.

Règle 5: Les valeurs des composantes R et G doivent être proches l'une de l'autre.

|R-G|<=15.

Règle 6: La valeur de la composante B doit être la plus petite.

R>B et G>B.

2.3.5.2 Algorithme C-Means flou et règles explicites

Un modèle basé sur la combinaison de la génération de règles floues à partir d'un algorithme de segmentation de type C-Means a été proposé par R. Vijayanandh et G. Balakrishnan en Janvier 2011 [213]. L'image à segmenter est d'abord transformée dans l'espace CIE L*a*b*, puis l'algorithme C-Means est appliqué afin d'effectuer un clustering sur les pixels de l'image. Le principe revient à minimiser la valeur J_A définie par l'équation suivante:

$$J_A = \sum_{i=1}^{C}\sum_{j=1}^{n}(u_{ij})^m (d_{ij})^2 \tag{2.71}$$

où u_{ij} correspond à la fonction d'appartenance du pixel j à la classe i, n représente le nombre de pixels et m une valeur fixé dans l'intervalle [1, ∞]. *C* est le nombre de clusters formés. Les contraintes suivantes sont définies sur u_{ij}.

$$\begin{cases} u_{ij} \in [0,1], & i = 1,2,\ldots,C \quad et \quad j = 1,2,..n \\ \sum_{i=1}^{C} u_{ij} = 1, & j = 1,2,\ldots,n \\ 0 < \sum_{j=1}^{n} u_{ij} < n, & i = 1,2,\ldots,C \end{cases} \tag{2.72}$$

$d_{ij}^2 = d^2(x_j, c_i)$ représente la distance euclidienne entre le centroide de la i^{ème} classe et le j^{ème} pixel. La solution de l'algorithme est itérative et se déroule comme suit:

1. Initialiser les centroides V= (c_1,c_2,…c_3)

2. La fonction d'appartenance est calculée comme suit:

$$u_{ij} = \frac{\left(\frac{1}{d^2(x_j, c_i)}\right)^{\frac{1}{m-1}}}{\sum_{i=1}^{C}\left(\frac{1}{d^2(x_j, c_i)}\right)^{\frac{1}{m-1}}}, \qquad j = 1,2,\ldots,n \tag{2.73}$$

3. mise à jour des centroides de V comme suit:

$$c_i = \frac{\sum_{j=1}^{N} (u_{ij})^m x_j}{\sum_{i=1}^{n} (u_{ij})^m}, \qquad i = 1,2,\ldots,n \tag{2.74}$$

4. Répéter jusqu'à ce que la valeur de J_A ne connaisse pas de significative variation.

Par la suite, les auteurs ont eu recours à des transformations de l'image en les espaces de couleur YCbCr et HSV et définissent des règles explicites dans ceux-ci afin de détecter les pixels correspondant à la peau. Ces règles se présentent selon les conditions suivantes:

$$\begin{cases} \max\{R,G,B\} - \min\{R,G,B\} > 15 \quad et \quad |R-G| > 15 \\ H \geq 0.0 \;\; et \;\; H \leq 0.20 \\ 77 \leq Cb \leq 127 \quad et \quad 122 \leq Cr \leq 173 \end{cases} \tag{2.75}$$

2.4 Discussion sur les performances des différentes méthodes de détections de la peau

Nous avons présenté dans ce chapitre, les approches de détection de la peau basées sur la couleur. On distingue dans les méthodes fondées sur ces approches, les méthodes paramétriques, les méthodes non paramétriques, les méthodes basées sur une fixation empirique des règles et des seuils de décision, les méthodes basées sur la logique floue ainsi que les méthodes hybrides.

Les méthodes basées sur la fixation empirique des règles et des seuils de décision sont des méthodes simples et rapides d'où leur large utilisation dans des applications qui fonctionnent en temps réel. Néanmoins, une reproche à ces méthodes est faite sur la façon dont les règles de décision et les seuils sont choisis. Le principal inconvénient de ce type de méthode est leur forte dépendance des conditions d'illumination. Les règles de seuillages ne sont efficacement valables que pour des conditions d'illumination précises. Un changement important sur ces conditions implique nécessairement de nouvelles règles à définir.

En s'appuyant sur quelques fonctions spécifiques paramétrées, les méthodes de peau paramétriques permettent d'ajuster les distributions colorimétriques des pixels appartenant aux images à traiter. La pertinence de ces méthodes et leur qualité de précision dépendent largement de la forme de distribution des pixels traités ainsi que de l'espace de couleur choisi. Cependant, les phases d'apprentissage et de test pour ce type de méthodes sont lentes puisque celles-ci impliquent une procédure d'évaluation des paramètres utilisés. Parmi les avantages de ces méthodes, on peut citer: gain en espace mémoire ainsi qu'en possibilité de manipulation, plus d'intelligence dans la régularité des distributions et capacité d'interpoler les données d'apprentissage quand elles sont dispersées.

Quand aux méthodes non paramétriques, elles sont généralement rapides dans la phase d'apprentissage et de test. Elles ne suggèrent aucune hypothèse sur la répartition des données d'apprentissage et elles sont indépendantes de la forme de distribution des pixels peau, contrairement aux méthodes paramétriques. Elles nécessitent cependant un espace de stockage important (pour représenter la carte de probabilité par exemple) ainsi qu'un ensemble d'apprentissage de grande taille. La phase d'apprentissage peut être plus longue avec les approches paramétriques, mais ces dernières possèdent l'avantage de fournir une bonne estimation de la densité avec un ensemble d'apprentissage plus réduit.

Les méthodes basées sur la logique floue sont moins efficaces que les autres types de méthodes et leurs performances se dégradent sous différentes conditions d'illumination [46]. Les méthodes hybrides semblent être beaucoup plus efficaces dans des conditions d'illuminations variées et dans des conditions de diversité ethnique de la couleur de peau [213], [225].

2.5 Conclusion

L'état de l'art montre qu'il existe dans la littérature de nombreuses méthodes pour la détection de régions de peau dans des images couleurs. Cinq types de méthodes se détachent principalement: les méthodes paramétriques, les méthodes non paramétriques, les méthodes explicites, les méthodes basées sur la logique floue et les méthodes hybrides. Les méthodes explicites sont basées sur une fixation empirique de règles et sont connues pour leur rapidité et simplicité. Les méthodes paramétriques se basent sur une estimation probabiliste des densités colorimétriques des pixels à traiter et suppose une forme analytique pour les distributions. Les méthodes non paramétriques ne dépendent d'aucune distribution et sont connus pour leur rapidité pendant la phase d'apprentissage et de test. Les chercheurs se sont intéressés récemment à proposer des méthodes hybrides combinant généralement un système de clustering flou avec un système basé sur la définition de règles explicites. Ces méthodes sont stables aux conditions d'illumination variées et d'appartenance ethnique de la couleur de peau. Nous avons opté pour une méthode non paramétrique pour la construction de notre système de détection de la peau. Dans le chapitre suivant, nous présenteront les méthodes d'apprentissages se basant sur l'estimation de paramètres les plus communément utilisées notamment celles qui ont une relation avec notre système.

CHAPITRE 3
METHODES D'APPRENTISSAGE BASEES SUR L'ESTIMATION DE PARAMETRES

3.1Introduction

L'utilisation montante des images numériques donne lieu à une multitude de nouvelles possibilités, comme la catégorisation automatique des photos, la détection de la peau humaine et la reconnaissance de visage. Ces problèmes doivent par essence être gérés informatiquement, mais peuvent être aisément résolus par un opérateur humain s'il dispose du temps nécessaire. Il est aujourd'hui possible d'effectuer ces tâches de manière automatique, après une phase *d'apprentissage* basée sur des observations passées. L'apprentissage vise à construire des hypothèses à partir d'exemples [96]. Simon [186] l'interprète comme les changements dans un système qui accomplira au mieux la même tâche, ou une tâche similaire dans la même population dans l'avenir. L'apprentissage statistique correspond au domaine se consacrant au développement d'algorithmes permettant à une machine d'apprendre à partir d'un ensemble de données, c'est-à-dire, d'y extraire des concepts et patrons caractérisant ces données. Bien que la motivation originale de ce domaine était de permettre la mise sur pied de systèmes manifestant une intelligence artificielle, les algorithmes issus de ce domaine sont maintenant répandus dans bien d'autres, tels la bioinformatique, la finance, la recherche d'information et l'imagerie. Plus spécifiquement, on peut définir un algorithme d'apprentissage comme suit: "*Un algorithme d'apprentissage est un algorithme prenant en entrée un ensemble de données D et retournant une fonction f* ". On désigne alors D comme *ensemble d'entraînement* ou *ensemble d'apprentissage* et la fonction f comme *modèle*. Suite à l'exécution d'un algorithme d'apprentissage, on dira que le modèle a été entraîné sur l'ensemble D [135].

Nous allons d'abord introduire les catégories de l'apprentissage statistiques (supervisé et non supervisé), les types d'approches discriminatives et génératives en ce qui concerne l'apprentissage supervisé et les types de méthodes d'apprentissage paramétriques et non paramétriques. Nous allons par la suite présenter un exemple d'apprentissage génératif avant de nous étaler sur la présentation des méthodes discriminatives sur lesquelles est fondé l'approche que nous avons adopté pour construire notre système.

3.2 Apprentissage statistique

L'apprentissage statistique appelé aussi *machine learning* se trouve au confluent de deux disciplines: l'intelligence artificielle et les statistiques. Il a comme but la résolution automatique de problèmes complexes à partir d'exemples. Son principe repose sur l'analyse d'exemples de réalisations d'un phénomène, cette analyse devant permettre de tirer des conclusions quand à la nature de la population dont ils sont issus et d'utiliser celles-ci pour comprendre le phénomène étudié ou classer de nouvelles réalisations. La démarche générale repose donc sur le principe d'induction à partir d'exemples. La croissance actuelle du nombre d'informations créées, stockées et manipulées chaque jour a favorisé la reconnaissance de l'apprentissage statistique comme l'une des sciences fondamentales de notre société de l'information. Les problèmes qu'elle aborde sont en nombre croissant et ils ne cessent de se diversifier [52]. La liste suivante donne quelques exemples d'applications:

- prédire si un patient risque une rechute au vu d'informations biologiques;
- ordonner les résultats d'une requête internet en fonction de leur pertinence;
- trouver un ensemble de pixels dans une image qui soient cohérents et qui délimitent des objets (segmentation d'images);
- trouver des archétypes de clients à partir d'analyse de tickets de caisses;
- reconnaître une adresse postale manuscrite automatiquement;

L'apprentissage statistique regroupe différentes méthodes visant à analyser des données en vue de mieux comprendre celles-ci ou de disposer d'une méthode de prédiction d'une ou de plusieurs variables d'intérêt. L'ensemble de données à analyser ainsi que l'objectif final de l'analyse peuvent prendre différentes formes. Cependant plusieurs problèmes de référence, d'intérêts pratique et théorique évidents, ont été formalisés et des solutions à ceux-ci ont été proposées. Pour une introduction générale au problème de l'apprentissage statistique, [68], [99], [211] et [19] constituent des ouvrages de référence en Anglais. En Français, il est possible de lire [178] et [90]. Nous présentons maintenant les deux contextes classiques de ce domaine, celui de l'apprentissage supervisé et celui de l'apprentissage non supervisé [233].

3.2.1 Apprentissage non-supervisé

L'apprentissage non supervisé rentre dans le cadre de la statistique dite descriptive. Son rôle est principalement *descriptif*. Il s'attache à isoler l'information utile au sein d'un jeu de données [22]. Il vise à trouver une structure cohérente au sein d'un ensemble de données susceptible d'en faciliter l'interprétation, l'analyse et la représentation. Il est utilisé lorsque l'on traite des données qui ne sont pas classés et dont on ne connait pas de classification. Il assure une organisation automatique des données en clusters. Parfois, il est possible de fixer le nombre de clusters, comme dans l'algorithme k-means [62].

L'apprentissage non-supervisé correspond au cas où aucune cible n'est prédéterminée. Ainsi, l'ensemble d'entraînement ne contient que des entrées:

$$D = \{X_t / X_t \in X\}_{t=1}^{n} \quad (3.1)$$

Ce type d'apprentissage ne définit pas explicitement la nature de la fonction *f* qui doit être retournée par l'algorithme d'apprentissage. Ainsi, c'est plutôt l'utilisateur qui doit spécifier le problème à résoudre. Par contre, dans tous les cas, le modèle capture certains éléments de la véritable distribution ayant généré *D* [135].

Supposons que nous disposons d'un algorithme d'apprentissage *A*. *A* produit une théorie issue de l'espace des hypothèses, que l'on peut utiliser pour expliquer la structure des données. Dans l'apprentissage non supervisé, l'algorithme *A* utilise un vecteur d'attributs $\vec{X} = (X_1, ..., X_p)$ pour essayer de trouver des *régularités* dans l'échantillon d'apprentissage. Elles se manifestent principalement par la constitution de groupes dans lesquels les observations diffèrent très peu au regard des valeurs prises par les X_i [96].

Les problèmes de référence de l'apprentissage non supervisé sont le problème du regroupement automatique et le problème de la réduction de la dimension, avec généralement pour objectif final la visualisation des données en deux dimensions. Nous allons présenter par la suite succinctement ces deux problématiques.

3.2.1.1 Le regroupement automatique

Le problème du regroupement automatique (*clustering*) [146] vise à trouver des groupes d'individus homogènes au sein des données. En d'autres termes, cela correspond à la recherche d'un partitionnement de l'espace X en g sous-ensembles (ou groupes) G_1, …, G_g. Chacun de ces sous-ensembles G_i peut facultativement être associé à un prototype X_i qui résume bien les données de *D* contenues dans G_i. Ainsi, *f* doit donner, pour chaque vecteur d'entrée, l'indice i du sous-ensemble G_i auquel il appartient. Le groupage peut aussi être utilisé afin de compresser un ensemble de données en le remplaçant par l'ensemble des prototypes. Il est même possible d'extraire des caractéristiques à l'aide du groupage, par exemple en considérant le vecteur binaire d'association aux différents sous-ensembles $(1_{f(x)=1}, \ldots, 1_{f(x)=g})$ comme vecteur de caractéristiques [135]. L'algorithme de groupage le plus simple et le plus répandu est certainement l'algorithme des K moyennes [140] [190] (*K-means algorithm*). Il s'agit d'un algorithme itératif, qui débute avec une initialisation aléatoire de la position des prototypes dans X, puis qui alterne entre (1) la réassignation de chaque vecteur d'entrée de *D* au prototype le plus proche selon la distance euclidienne et (2) la mise à jour des prototypes à la valeur de la moyenne des vecteurs d'entrée associés au même prototype. La figure 3.1 montre quelques étapes de l'exécution de cet algorithme sur un ensemble d'entrées en 2 dimensions. Les cercles représentent les entrées et les carrés représentent les prototypes. De

gauche à droite, l'état de l'algorithme est illustré à l'initialisation, après une étape et après 3 étapes.

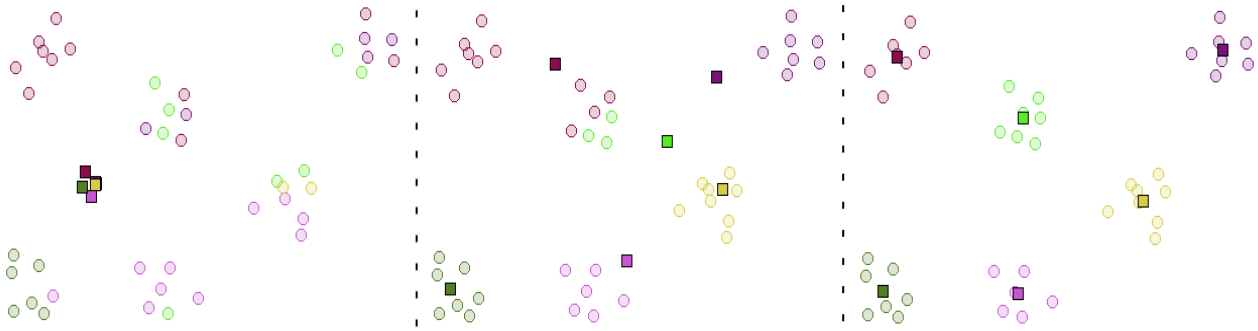


Figure. 3.1: Exemple de groupage obtenu à l'aide de l'algorithme des K moyennes [135].

3.2.1.2 La réduction de la dimension

Le problème de réduction de la dimension vise à extraire du jeu de données initial de nouvelles variables peu nombreuses portant le plus d'information possible sur le jeu de données initial et permettant par exemple la représentation en deux dimension du jeu de données. En d'autres termes, Pour cette tâche, f doit associer à un vecteur d'entrée x une représentation f(x) de plus petite dimensionnalité que x mais conservant l'essentiel de l'information contenue dans l'entrée. C'est ici que nous retrouvons les méthodes telles que l'analyse en composantes principales et l'analyse en composantes indépendantes. La première de ces méthodes est aussi la plus ancienne [160] [105]. Cette méthode linéaire très largement utilisée permet de réduire la dimension de l'espace de travail. L'analyse en composantes indépendantes, quant à elle, vise à extraire de nouvelles variables qui soient statistiquement indépendantes entre elles [52].

3.2.2 Apprentissage supervisé

L'apprentissage supervisé a un rôle *prédictif*: Il permet d'évaluer la distribution d'une quantité (eg. la taille d'un individu) sans la mesurer directement, mais en se basant sur des valeurs qui lui sont liées (eg. le poids de la personne) [22]. Il correspond au cas où l'objectif de l'apprentissage est déterminé explicitement via la définition d'une cible à prédire. Dans ce cas, *D* correspond à un ensemble de n paires d'*entrées* x_t et de *cibles* associées y_t :

$$D = \{(x_t, y_t) / x_t \in X,\ y_t \in Y\}_{t=1}^{n} \qquad (3.2)$$

Typiquement, un tel ensemble est récolté en fournissant l'ensemble des entrées à un groupe de personnes et en leur demandant d'associer à chacune de ces entrées une cible appropriée dans le contexte du problème à résoudre. La tâche d'un algorithme d'apprentissage est alors d'entraîner un modèle qui puisse imiter ce processus d'étiquetage par un humain, i.e. qui puisse prédire pour une entrée x quelconque la valeur de la cible y qui aurait normalement été donnée par un humain. Cependant, les algorithmes d'apprentissage ne se limitent pas à la modélisation du comportement de l'humain et peuvent être utilisés pour modéliser la relation liant des paires d'entrées et de cibles provenant d'un autre phénomène (e.g., la relation entre une action et son prix à la bourse telle que générée par les marchés financiers). La nature de l'ensemble Y d'où proviennent les cibles dépendra du type de problème à résoudre [135].

L'apprentissage supervisé vise toujours à partir d'un vecteur d'attributs $\vec{X}$ que l'on nomme ici attributs prédictifs, ou encore variables exogènes de construire une fonction ou concept sous-jacent *f* tel que:

$$Y = f(\vec{X}) \tag{3.3}$$

Y est qualifiée de variable à prédire, ou encore de variable endogène.

L'apprentissage permet de mettre à jour la fonction *f* que l'on nomme classifieur ou prédicteur avec comme objectif $\hat{Y} = Y$ où $\hat{Y}$ représente la véritable variable de sortie correspondant à l'individu traité. La modélisation du lien entre *X* et *Y* est donc primordiale.

Selon la nature de Y, nous distinguons généralement deux familles d'apprentissage supervisé: La *régression* et la *classification* [96].

3.2.2.1 Régression

Dans un problème de régression, Y correspond à un ensemble de valeurs continues ou de vecteurs de valeurs continues. Voici quelques exemples de tels problèmes:

Prédiction du poids: x ∈ X correspond à des caractéristiques physiques d'une personne (par exemple son âge, son sexe, etc.) et y ∈ Y correspond à son poids.

Prédiction de la valeur d'une action: x ∈ X correspond à une séquence d'indicateurs du marché pour différentes journées et y ∈ Y correspond à la progression de la valeur d'une action [135].

3.2.2.2 Classification

3.2.2.2.1 Notion de classe

Traditionnellement, la notion de classe dans la classification a souvent été synonyme de *thème*, on parle alors de classification thématique. La problématique revient donc à partitionner un ensemble d'objets autour de thématiques que l'on

désigne communément par *classes*. Dans ce cas, il s'agit par exemple, de savoir si un document concerne le *thème botanique*, le *thème philosophie* ou bien le *thème physique*. Cependant, la tâche de classification a évolué au même temps que les besoins. Cette évolution a engendré de nouvelles problématiques pour lesquelles les classes ne correspondent plus nécessairement à des thématiques. A titre d'exemple, nous pouvons citer le problème de filtrage qui est définis comme étant un processus qui permet d'extraire à partir d'un flot d'informations (news, emails, actualités journalières) celles qui sont susceptibles d'intéresser un utilisateur. Voici quelques autres exemples:

- Un logiciel qui détecte si un message électronique correspond à un spam ou non.

- Un système qui sélectionne des informations pertinentes dans des flux de dépêches et les envoie à différents organismes concernés.

D'autres problématiques émergentes telles que la le tri de courrier électronique dans différentes boites aux lettres personnelles (mails urgents, mails privés, mails du DG) et la distinction entre les pixels peau et non peau dans l'analyse d'images ou de séquences vidéos ne peuvent être qualifiée de problématiques thématiques [62].

3.2.2.2.2 Principe de la classification

Dans le cadre d'un problème de classification, Y correspond à un ensemble fini de classes (Y est discrète) auxquelles peuvent appartenir les différentes entrées possibles $x \in X$. Voici d'ailleurs quelques exemples de problèmes de classification:

Reconnaissance de caractères: $x \in X$ correspond à la représentation vectorielle de la valeur des pixels d'une image et $y \in Y$ au caractère associé (par exemple, un chiffre entre 0 et 9).

Détection de la peau humaine: $x \in X$ correspond à la représentation vectorielle de la valeur des pixels d'une image et $y \in Y$ à la classe associée (peau ou non peau).

Le problème de classification est le thème principal qui nous intéresse. Certains auteurs insistent sur la différence entre *régression* et *classification* en précisant que la classification est un problème plus simple que la régression [169][110]. L'une des finalités de l'apprentissage supervisé est la prévision. Prenons l'exemple d'identification des pixels de peau dans une image. La variable à prédire ici est la variable *classe du pixel*, qui ne prend que deux modalités *peau* ou *non peau*. Pour des raisons diverses telles les conditions d'éclairage, la diversité ethnique etc., l'identification du pixel de peau est un problème complexe. C'est la raison pour laquelle nous cherchons un moyen pour prédire la classe du pixel (peau /non peau).

3.2.2.2.3 Contextes de classification

Les contextes sont des facteurs prépondérants qui ont une influence directe sur les modèles utilisés ainsi que sur la façon de les utiliser. Il existe différents contextes de classification. Nous allons définir les plus connus et utilisés d'entre eux.

3.2.2.2.3.1 Classification bi classes

La classification bi classe est la problématique pour laquelle le système de classification répond à la question "*L'objet est-il pertinent ou non pour une classe donnée*". Par exemple, un message est-il un spam ou non ? Cette problématique est le contexte que l'on désigne d'ordinaire de filtrage. C'est dans le cadre de ce contexte que s'inscrit le domaine de la détection de la peau par une approche couleur. Il s'agit de déterminer si un pixel donné appartient à la classe peau ou non [62].

3.2.2.2.3.2 Classification multi classes disjointes

La classification multi classes disjointes correspond au contexte où nous sommes en présence de plusieurs classes (>1) et pour lequel un objet appartient *exactement* à une seule classe. La classification à classes disjointes est la problématique pour laquelle le système de classification répond à la question "*A quelle classe (au singulier) appartient tel objet ?*" [62].

3.2.2.2.3.3 Classification multi classes

C'est le cas le plus général de la classification. Dans cette problématique, le système de classification associe zéro (0), une (1) ou plusieurs (> 1) classes à un objet. Autrement dit, le système doit répondre à la question "*Quelles sont les classes (au pluriel) auxquelles appartient l'objet ?*". Dans ce cas, il n'est pas difficile de constater qu'un problème de n classes se résout comme n problème bi classes [62].

Nous pouvons différencier deux types d'approches pour résoudre un problème d'apprentissage supervisé: approches discriminatives et approches génératives. Nous nous intéresserons dans la section suivante à présenter ces deux types d'approches.

3.3 Apprentissage discriminatif et génératif

Dans beaucoup d'application d'apprentissage statistique, l'objectif est d'assigner à une observation $x \in X$ une classe $y \in Y$. L'estimation est effectuée par une fonction de décision paramétrée. Les paramètres sont appris par le système à partir d'un ensemble d'apprentissage constitué de données d'entrée $X = \{x_1, \ldots, x_N\}$ et de données de sortie $Y = \{y_1, \ldots y_N\}$. Les approches d'apprentissages discriminatives et génératives utilisées de nos jours sont opposées les unes aux autres [188]. Par exemple, dans [113], elles sont même présentées comme différentes écoles de pensée distinctes.

3.3.1 Approche discriminative

Les approches discriminatives ou de régression consiste à modéliser les relations qui lient les entrées aux sorties du système, avec un minimum d'hypothèses sur la structure des données d'entrée. Ces méthodes répondent directement à l'objectif de l'utilisateur en se focalisant sur la règle de décision plutôt que sur l'interprétation de cette décision. Cette approche permet d'estimer une fonction avec un minimum

d'hypothèses sur la forme de la distribution des entrées et permet de modéliser directement la règle de classification *P*(*Y* |*X*). Les paramètres (ou les coefficients dans le cas des méthodes non-paramétriques) de ce modèle sont estimés par minimisation du coût de classification [22]. La minimisation du coût est équivalente au maximum de la fonction de vraisemblance dont nous allons présenter le principe ultérieurement.

3.3.2 Approche générative

Les classifieurs basés sur les approches génératives (parfois appelées informatives [175]) portent différents noms dans la littérature. En 1996, Ripley parle de *plug-in rule classifier* [171]. Plusieurs auteurs utilisent le terme *Bayesian Classifiers* [67] [65] [134]. Les approches *génératives* modélisent en premier lieu la structure du système [22]. Cela consiste à trouver une structure pour la distribution jointe des entrées *X* et sorties *Y P(X, Y).* Ensuite, l'application de la loi de Bayes permet d'en déduire la solution recherchée *P(Y |X)* [52].

Ce type d'approche correspond à une vaste catégorie de classifieurs, dont les plus simples (et souvent très performants) sont l'*Analyse Discriminante Linéaire* (*LDA*) et les *classifieurs de Bayes Naïfs* (NB). On parle de *modélisation générative* car une loi de probabilité jointe est définie pour *toutes* les variables possibles, c'est-à-dire pour les données à prédire *Y*, les données d'entrée *X et* des variables annexes *Z* non observées (cachées ou latentes) [22].

Par rapport aux méthodes discriminatives qui ne s'intéressent qu'aux relations entrées-sorties, ces approches modélisent en plus les liens qui existent au sein des variables d'entrée. Ainsi, le travail de modélisation est plus important. La modélisation générative, parfois qualifiée d'approche *informative*, tient son nom de l'aspect explicatif qui la lie aux données. Elle répond à la question "Comment pourrait-on générer les données que l'on observe ?" [22].

3.3.3 Comparaison des méthodes génératives et discriminatives

En apprentissage supervisé, les méthodes génératives sont parfois considérées comme sous-optimales. Une citation écrite par Vapnik en 1998 à propos de ces méthodes résume bien cette pensée : "*One should solve the* [classification] *problem directly and never solve a more general problem as an intermediate step* [comme modéliser la distribution de toutes les données]" [209]). Cette remarque a probablement freiné le développement des méthodes génératives au profit des méthodes discriminatives qui résolvent le problème *directement*. Les méthodes génératives sont connues pour leur précision à trouver la probabilité à postériori correcte à condition que les données d'apprentissage soient définies dans la distribution adéquate [22]. Cependant, dans des applications réelles, la véritable distribution des données est inconnue. Il est alors dans ce cas inutile de construire la distribution jointe des variables exogènes avec les variables endogènes si la probabilité conditionnelle P(Y/X) peut être modélisée directement. Cependant, lors de la construction des intervalles définissant les classes comme mentionné dans [206],

le système d'apprentissage employé par une méthode discriminative nécessite un entrainement sur tout type de données appartenant à toutes les classes incluant les classes positives (sur lesquelles l'intérêt se porte) et négatives (toutes les autres). Dans le cas de la détection de la peau par exemple, ces méthodes effectuent un apprentissage à la fois sur la classe peau (classe positive) et la classe non peau (classe négative). Dans [163], les auteurs insistent sur le fait que les algorithmes de type discriminatifs résolvent des problèmes d'apprentissage que les algorithmes de types génératifs n'arrivent pas à résoudre. Toutefois, les données appartenant aux classes négatives ne sont pas toujours disponibles et cela peut constituer un inconvénient majeur pour les méthodes d'apprentissage discriminatives.

Ces deux approches sont distinctes et complémentaires. Lorsqu'un grand nombre de données d'apprentissage sont disponibles, certaines méthodes discriminatives non linéaires sont très performantes. A l'inverse, les méthodes génératives offrent des outils de modélisation élaborés qui permettent de limiter la quantité de données nécessaires à l'apprentissage. Ce type d'approche générative cherche à modéliser la distribution jointe $P(X, Y)$ des entrées et des sorties. On en déduit ensuite la règle de classification par application de la loi de Bayes [175] [22].

3.3.4 La statistique bayésienne

La statistique bayésienne consiste à considérer que les paramètres d'un modèle sont aléatoires. Il est donc nécessaire de définir une distribution *à priori* de ces paramètres ne dépendant pas des observations. Autrefois très contestée pour le caractère subjectif de cette loi *à priori*, l'approche bayésienne est aujourd'hui un outil reconnu de l'apprentissage statistique. Elle permet d'incorporer de manière probabiliste des informations extérieures aux données, souvent cruciales pour la reconnaissance.

Les approches génératives et discriminatives dans un cadre bayésien sont représentées sous la forme de modèles graphiques sur la figure 3.2. L'approche générative modélise la densité jointe des entrées et des sorties alors que l'approche discriminative définit une distribution des sorties conditionnellement aux entrées et aux paramètres. Dans la majorité des cas, la distribution *à priori* des paramètres ne dépend pas des entrées, ce qui explique l'absence de lien entre X et θ [22].

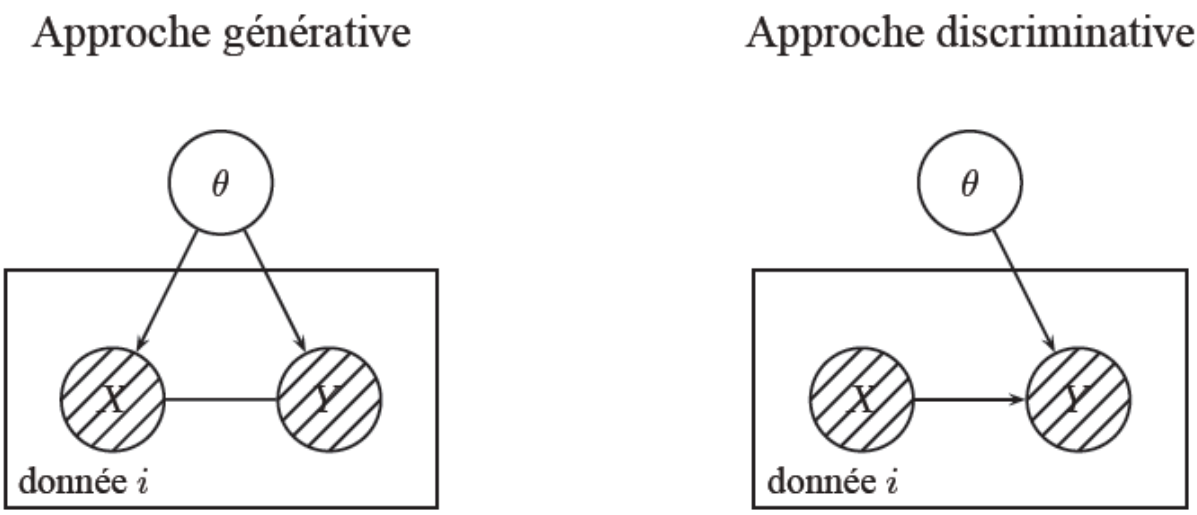


Figure. 3.2: Modèles graphiques illustrant les approches bayésiennes pour la classification supervisée [22].

X représente les entrées (ou covariables), *Y* les sorties à prédire et θ le vecteur des paramètres. Les variables observées sont indiquées par des hachures. Les cadres (*plates* en anglais) indiquent une répétition indépendante des variables.

3.4 Modèles paramétriques et non paramétriques

La définition de modèles de densités (jointes ou conditionnelles) pour les variables considérées s'avère nécessaire pour effectuer l'apprentissage statistique. Souvent, Dans de nombreuses applications, on ne dispose pas directement d'un modèle adapté à des données multidimensionnelles. Le choix du modèle de densité est bien entendu très important pour obtenir un taux de classification convenable. On peut modéliser ces densités par un *modèle paramétrique* ou un modèle *non paramétrique*. Le principe d'un modèle paramétrique consiste à contraindre les distributions à épouser un modèle rigide comportant un nombre fini de paramètres, c'est-à-dire que la famille de probabilités qu'il définit peut être paramétrée par un vecteur φ dans un espace Φ de dimension finie. Lorsqu'on a une idée précise de la structure des données, il est généralement aisé de paramétrer la distribution.

En revanche, les méthodes non paramétriques considèrent un espace paramétrique dont la dimension dépend de l'échantillon d'apprentissage et toute probabilité peut être approchée par une fonction de la classe considérée. Les modèles non-paramétriques, au contraire, deviennent de plus en plus complexes au fur et à mesure que le nombre d'exemples d'apprentissage augmente. Leur capacité à représenter des phénomènes complexes progresse donc avec ce nombre. Ces modèles non paramétriques sont efficaces dans les méthodes discriminatives car ils se basent sur un minimum d'hypothèses sur la frontière de classification ou la fonction à prédire. Notons que ces méthodes considèrent souvent une base de

fonctions (noyaux, splines, Fourier), ce qui permet de percevoir un problème de classification non linéaire dans l'espace d'origine de manière linéaire dans l'espace fonctionnel (communément appelé *feature space*) [22]. Dans le cadre de notre travail, nous étudierons un estimateur non paramétrique à noyaux.

3.5 Modèles de densité et règles de classification

On cherche à prédire au mieux la valeur d'une variable $y \in Y$ en fonction de variables associées $x \in X$ (aussi appelées covariables), où X et Y sont des espaces mesurables quelconques. D'un point de vue probabiliste, notre objectif est d'estimer la distribution conditionnelle $P(Y|X)$. Ne connaissant pas les valeurs des variables X qui seront utilisées dans le futur, nous supposons qu'elles suivent une distribution $P(X)$. Cela revient à considérer que la distribution jointe des données $P(X, Y) = P(X) P(Y|X)$ existe.

3.5.1 Le modèle

Comme nous l'avons déjà précisé, le parti pris de la classification générative est de modéliser la densité jointe P(X, Y) par un modèle P(X, Y; θ) paramétré par $\theta \in \Theta$, puis de prédire les données de test en appliquant la règle de Bayes pour trouver p(Y |X). L'approche générative se définit comme suit:

Soit p(X, Y; θ), $\theta \in \Theta$ un modèle pour la densité jointe de X et de Y définie sur X x Y. Une méthode d'apprentissage génératif est définie par:

1. une phase d'apprentissage: recherche d'un estimateur $\hat{\theta}$ de θ à partir des données d'apprentissage $(x, y) = \{(x_i, y_i), i = 1, \ldots, n\}$.

2. une phase de test : à partir d'une donnée X, la valeur y est prédite avec une probabilité:

$$P(Y = y \mid X = x; \hat{\theta}) = \frac{P(X = x, Y = y; \hat{\theta})}{\int_y P(X = x, Y = y; \hat{\theta}) dy} \tag{3.4}$$

Dans le cadre de la *classification supervisée*, la variable à prédire Y est discrète et définie l'index de la classe. La formule (3.4) peut s'écrire comme suit:

$$P(Y = k \mid X = x; \hat{\theta}) = \frac{\pi_k f_k(x; \hat{\theta})}{\sum_{l=1}^{K} \pi_l f_l(x\ ; \hat{\theta})} \tag{3.5}$$

pour $k = 1, \ldots, K$, où $\pi_k = P(Y = k)$ et f_k est la densité de la $k^{\text{ième}}$ classe [22].

3.5.2 L'apprentissage des paramètres

La notion d'apprentissage se traduit ainsi par l'estimation du paramètres θ de manière à satisfaire au mieux les objectifs du modélisateur, à savoir approcher par le

modèle $p(Y|X; \theta)$ la distribution conditionnelle des données $p(Y|X)$. Classiquement, les estimateurs sont solutions d'un problème d'optimisation, généralement sous la forme d'une minimisation d'une fonction de coût dépendant des données d'apprentissage labellisées (x, y) et du vecteur de paramètres θ.

Il est possible de maximiser deux types de vraisemblance:

- La vraisemblance jointe des données:

$$L_J(\theta) = \sum_{i=1}^{n} \log P(x_i, y_i; \theta) \tag{3.6}$$

est souvent utilisée car c'est la vraisemblance *naturelle* du modèle paramétrique. Elle donne lieu à l'estimateur *génératif* (parfois appelé *informatif* [175]):

$$\hat{\theta}_J = \arg\max_{\theta \in \Theta} L_J(\theta) \tag{3.7}$$

- La vraisemblance conditionnelle des données:

$$L_C(\theta) = \sum_{i=1}^{n} \log P(y_i | x_i; \theta) \tag{3.8}$$

La quantité $L_C(\theta)$ est aussi appelée *entropie de classification*. Elle mesure la faculté d'un modèle à séparer les données. Elle donne lieu à l'estimateur *discriminatif*:

$$\hat{\theta}_C = \arg\max_{\theta \in \Theta} L_C(\theta) \tag{3.9}$$

Afin d'affermir notre compréhension sur les types de méthodes d'apprentissage supervisé, nous allons présenter dans la section suivante un modèle d'apprentissage non supervisé basé sur la loi de Dirichlet avant d'aborder dans la section 3.7 un exemple célèbre de méthodes de types générative qui est la régression linéaire discriminante. La section 3.8 fera objet de présentation d'une méthode de type discriminative qui constitue une partie de la base théorique sur laquelle est fondée l'approche que nous avons adoptée. Il s'agit de la régression logistique et ses variantes.

3.6 Modèle d'apprentissage non supervisé basé sur la loi de Dirichlet

La mixture de distribution beta correspondant à la loi de Dirichlet a été proposée par N. Bouguilla et D. Ziou [24] en 2007 dans le cadre de la détection de la peau humaine. Une revue de cette méthode a été effectué par S. Boutemdjet avec les auteurs que nous venons de citer [200] en 2009 dans le cadre de la catégorisation d'images. Il s'agit d'une méthode d'apprentissage non supervisée où le nombre de classes à trouver est inconnu à priori. Considérons un espace de dimension d. La

fonction de probabilité de la densité donnée par la loi généralisée de Dirichlet est donnée par la fonction suivante:

$$P(X_1,...,X_d) = \prod_{i=1}^{d} \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} X_i^{\alpha_i - 1} \left(1 - \sum_{j=1}^{i} X_j\right)^{\gamma_i}$$

(3.10)

avec $\sum_{i=1}^{d} X_i < 1$, $0 < X_i < 1$ pour i=1…d, $\alpha_i > 0$, $\beta_i > 0$, $\gamma_i = \beta_i - \alpha_{i+1} - \beta_{i+1}$ pour i=1…d-1 et $\gamma_d = \beta_d - 1$.

Lorsque $\beta_i = \alpha_{i+1} + \beta_{i+1}$, l'équation (3.10) peut s'écrire de la manière suivante:

$$P(X_1,...,X_d) = \frac{\Gamma(\alpha_1 + \alpha_2 + ... + \alpha_d + \alpha_{d+1})}{\Gamma(\alpha_1)\Gamma(\alpha_2)... \Gamma(\alpha_d)\Gamma(\alpha_{d+1})} \left(1 - \sum_{i=1}^{d} X_i\right)^{\alpha_{d+1} - 1} \prod_{i=1}^{d} X_i^{\alpha_{d-1}} \quad (3.11)$$

Avec $\alpha_{d+1} = \beta_d$.

Les moyennes et variances de la distribution de Dirichlet se présentent comme suit:

$$E(X_i) = \frac{\alpha_i(\beta_i + 1)}{\alpha_i + \beta_i \prod_{k=1}^{i-1} (\alpha_k + \beta_k)} \quad (3.12)$$

$$Var(X_i) = E(X_i)\left(\frac{\alpha_i + 1}{\alpha_i + \beta_i + 1} \prod_{k=1}^{i-1} \frac{\beta_k + 1}{\alpha_k + \beta_k + 1} - E(X_i)\right) \quad (3.13)$$

La covariance entre deux éléments X_i et X_j est donnée par :

$$\mathrm{cov}(X_i, X_j) = E(X_j)\left(\frac{\alpha_i}{\alpha_i + \beta_i + 1} \prod_{k=1}^{i-1} \frac{\beta_k + 1}{\alpha_k \beta_k + 1} - E(X_i)\right) \quad (3.14)$$

La loi de Dirichlet avec M composantes (clusters) est définie comme suit:

$$P(\vec{X} / \Theta) = \sum_{j=1}^{M} P(\vec{X} / \alpha_j) P(j) \quad (3.15)$$

Pour $\vec{X}_i \in \vec{X}$, nous avons

$$P(\vec{X}_i / \Theta) = \sum_{j=1}^{M} P(j) \prod_{l=1}^{d} P(X_{il} / \alpha_{jl}, \beta_{jl}) \quad (3.16)$$

Avec 0< P(j)< 1 et $\sum_{j=1}^{M} P(j) = 1$. Dans ce cas, les paramètres de la mixture pour les M clusters est donné par $\Theta = (\alpha, \vec{P})$ avec $\alpha = (\vec{\alpha}_1,....,\vec{\alpha}_M)^T$, $\vec{\alpha}_j = (\alpha_{j1}, \beta_{j1},...., \alpha_{jd}, \beta_{jd})$ pour j=1,…, M et $\vec{P}(P(1),...,P(M))^T$ représente le vecteur des probabilités à priori des M clusters.

Soit $\vec{Z}_i = (Z_{i1}, Z_{i2},....,Z_{im})$ un vecteur associé à l'individu $\vec{X}_i$ où $Z_{ij} \in \{0,1\}$ (j =1…m) une variable binaire indiquant l'appartenance d'un individu $\vec{X}_i$ à la classe j. La distribution de $\vec{X}_i$ étant donnée par l'équation suivante:

$$P(\vec{X}_i / Z_i, \Theta) = \prod_{j=1}^{M} \left(\prod_{l=1}^{d} P\left(X_{il} / \alpha_{jl}, \beta_{jl}\right) \right)^{Z_{ij}} \quad (3.17)$$

Les M clusters sont définis par application de l'algorithme C-means présenté précédemment au chapitre 2, section 2.4.3.3.2. L'estimation optimale des paramètres de la distribution de Dirichlet se fait par minimisation de la valeur V définie par l'équation suivante:

$$V \approx -\log(h(\Theta)) - \log\left(P(X / \Theta) + 0.5\log(|F(\Theta)|)\right) + \frac{N_p}{2}(1 + \log(k_{Np})) \quad (3.18)$$

avec N_p représente le nombre de paramètres à estimer et k_{Np} une constante.

$|F(\Theta)| \approx |F(\vec{P})| \prod_{j=1}^{M} |F(\vec{\alpha}_j)|$ représente la matrice d'information de Fisher.

$|F(\vec{P})|$ est défini comme suit:

$$|F(\vec{P})| = \frac{{N_p}^{M-1}}{\prod_{j=1}^{M} P(j)} \quad (3.19)$$

L'information de Fisher par rapport au vecteur $\vec{\alpha}_j$ est définie comme suit:

$$|F(\vec{\alpha}_j)| \approx \prod_{l=1}^{d} |F(\alpha_{j1}, \beta_{j1})| \quad (3.20)$$

Avec $|F(\alpha_{j1}, \beta_{j1})|$ représente l'information de Fisher de la distribution Beta avec les paramètres (α_{j1} , β_{j1}). La distribution $h(\Theta)$ est définie comme suit:

$$h(\Theta) = h(\vec{P})h(\alpha) \tag{3.21}$$

Avec $\vec{P}$ est défini sur l'ensemble suivant: $\left\{(P(1),....,P(M)): \sum_{j=1}^{M} P(j) = 1\right\}$.

$$h(\vec{P}) = \frac{\Gamma(\sum_{j=1}^{M} n_j)}{\prod_{j=1}^{M} \Gamma(n_j)} \prod_{j=1}^{M} P(j)^{n_j - 1} \tag{3.22}$$

Avec $\vec{n} = (n_1,..., n_{2d})$ est le vecteur des paramètres de la distribution de Dirichlet.

$$h(\alpha) = \prod_{j=1}^{M} h(\vec{\alpha}_j) \tag{3.23}$$

avec:

$$h(\vec{\alpha}_j) = \frac{\Gamma\left(\sum_{l=1}^{2d+1} n_l\right)}{(2de^5)^{\sum_{l=1}^{2d+1} n_l - 1} \prod_{l=1}^{2d+1} \Gamma(n_l))} \left(2de^5 - \sum(\alpha_{j1} + \beta_{j1})\right)^{n_{2d+1}-1} \prod_{l=0}^{d-1} \alpha_{jl+1}^{n_{2l+1}-1} \beta_{jl+1}^{n_{2l+2}-1} \tag{3.24}$$

Avec $\vec{\alpha}_j$ est défini dans l'ensemble $\left\{(\alpha_{j1}, \beta_{j1},....,\alpha_{jd}, \beta_{jd}): \sum_{l=1}^{d}(\alpha_{jl} + \beta_{jl}) < 2de^5\right\}$

Dans [200], S.Boutemdjet, N.Bouguila et D.Ziou ont proposés un modèle basé sur une mixture de K distributions beta notée $\square_l$={ $\square_{1l}$,$\square_{2l}$ …, $\square_{kl}$}. Φ_{il} représente une variable binaire ajustée à 0 quand X_{il} est issu de D_l et 1 sinon avec D_l la distribution commune à toutes les données.

Soit le vecteur $\vec{W}_i = (\vec{W}_{i1},...,\vec{W}_{id})$ où $\vec{W}_{il} = (W_{il1},...,W_{ilk}) \in \{0,1\}^k : \sum_k W_{ilk} = 1$. Avec W_{ilk}=1 si X_{il} est issu du k^{ième} compsante $\square_{kl}$ avec ϕ_{il}=0. Nous avons alors:

$$P(\vec{X}_i / \vec{Z}_i, \Theta) \approx P(\vec{X}_i / \vec{Z}_i, \vec{\phi}_i, \vec{W}_i) = \prod_{j=1}^{M} \left[\prod_{l=1}^{d} P(X_{il} / \theta_{jl})^{\phi_{il}} \left(\prod_{k=1}^{K} P(X_{il} / \varepsilon_{kl})^{W_{ilk}} \right)^{1-\phi_{il}} \right]^{Z_{ij}} \tag{3.25}$$

L'estimation optimale des paramètres de la distribution de Dirichlet se fait cette fois-ci par minimisation de la valeur V définie par l'équation suivante:

$$V = -\log h(\Theta) + 0.5\log\left|F(\Theta)\right| + \frac{c}{2}\left(1+\log\frac{1}{12}\right) - \log P(X/\Theta) \quad (3.26)$$

Avec *c* représente une constante correspondant au nombre de paramètre du modèle.

La probabilité à postériori correspondant à l'appartenance de l'individu $\vec{X}_i$ étudié à l'une des classes considérées j est défini comme suit:

$$\hat{Z}_{ij} = \frac{P(\vec{X}_i/\vec{\Theta}_j)P(j)}{\sum_{j=1}^{M} P(\vec{X}_i/\vec{\Theta}_j)P(j)} \quad (3.27)$$

3.7 La régression linéaire discriminante

Introduite en 1936 par Fisher, l'Analyse Linéaire Discriminante (*Linear discriminent analysis* ou *LDA*) est probablement la méthode générative la plus ancienne et la plus populaire [74].

LDA modélise la densité de chaque groupe par une distribution gaussienne multivariée, avec une matrice de covariance commune. On en déduit une règle de classification par la formule de Bayes (3.5). C'est donc une méthode générative pour laquelle la densité des classes s'écrit:

$$f_k(x;\theta) = \frac{1}{\sqrt{2\pi}^d \left|\sum\right|^{0.5}} \exp\left\{\frac{1}{2}(x-m_k)^T \sum^{-1}(x-m_k)\right\} \quad (3.28)$$

Où θ={m1,…,mk, $\sum^{-1}$} représente les paramètres à estimer. m_k correspond à la moyenne de la classe k et ∑ à la matrice de covariance.

La frontière de classification *P*(*Y* |*X* = *x*; *θ*) résultante peut avoir la forme logistique suivante:

$$P(Y=1|X=x;w) = \frac{1}{1+e^{-w_{1:d}x - w_0}} \quad (3.29)$$

Les paramètres w_i sont estimés comme suit :

$$\begin{cases} w_{1:d} = \sum^{-1}(m_1 - m_2) \\ w_0 = \left(\frac{m_1+m_2}{2}\right)' \sum^{-1}(m_1 - m_2) \end{cases} \quad (3.30)$$

L'équation (3.29) est également employée par La régression logistique linéaire. C'est la méthode de type discriminative de référence qui modélise directement la distribution conditionnelle de la variable *Y* sachant le prédicteur $X \in R^d$ par une loi binomiale dont le paramètre dépend de manière logistique des covariables. Nous allons présenter ce type de méthode ultérieurement [22].

L'analyse discriminante linéaire est concernée par les problèmes de classification où les variables dépendantes sont catégoriques (nominales ou ordinales) et les variables indépendantes représentent des métriques. Son principe est d'expliquer et de prédire l'appartenance d'un individu à une classe (groupe) prédéfinie à partir de ses caractéristiques mesurées à l'aide de variables prédictives (indépendantes). Cela revient à construire une fonction discriminante linéaire qui sépare les données de chaque classe à part. Cette fonction agit sur une droite à définir d'une manière optimale (*droite de projection*) de façon à séparer les projections des données des deux classes considérées sur elle [121]. L'équation de la droite discriminante dans un espace multi dimensionnel est défini comme suit:

$$Z = w_0 + \sum_{i=1}^{d} w_i X_i \tag{3.31}$$

Les paramètres w_0 et w_i sont définis par l'équation (3.30) avec m_1 et m_2 sont les projections des moyennes sur la droite de projection. Celle-ci est calculée de manière à discriminer les deux classes considérées de la meilleure façon.

X_i correspond à la $i^{ème}$ dimension de l'espace et w_i est un poids associé à X_i. Z est appelé le *score de discrimination*.

L'interprétation géométrique du score de discrimination dans un espace à deux dimensions est illustrée sur la figure suivante:

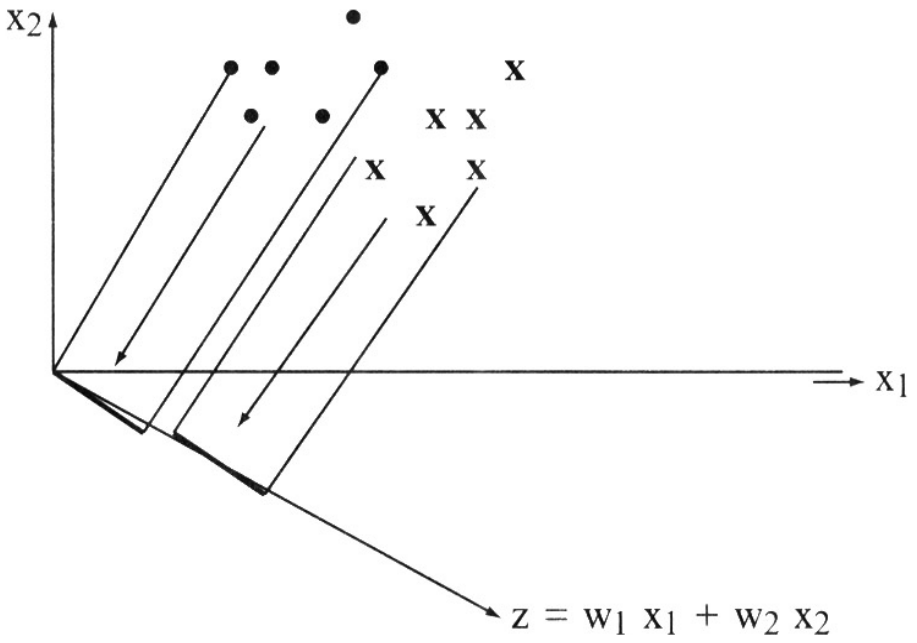


Figure. 3.3: Interprétation géométrique de la régression linéaire discriminante dans un espace 2D [121].

Comme montrée sur la figure 3.3, le score de discrimination pour un individu traité correspond à la projection du point qui lui est associé dans l'espace de représentation sur la droite discriminante: $Z= w_1 X_1+w_2 X_2$. Dans ce cas, nous considérons $w_0 =0$.

La construction de la droite discriminante nécessite un ajustement des paramètres w_i qui permet de maximiser la variance inter-classes et de minimiser la variance intra-classes du score de discrimination de l'échantillon de données qui sert d'apprentissage et dont on connait préalablement la classification. Une fois construite, la fonction de discrimination Z est utilisée pour prédire la classe de nouvelles données non classifiées. Le *score de coupe (cutting score)* est utilisé comme critère de classification de nouvelles données à traiter. Il constitue un seuil de séparation entre deux classes prédéfinies. Le choix du score de coupe dépend de la distribution des données d'apprentissage dans les classes considérées. Soient Z_a et Z_b les moyennes des scores de discrimination des échantillons de données préalablement classifiés dans les classes A et B respectivement. Considérons le cas où lors de la phase d'apprentissage, les données sont réparties d'une manière équilibrée entre les deux classes prédéfinies (c'est-à-dire que celles-ci sont de même taille et que les variances des données leurs correspondant sont uniformes). Le choix optimal du score de coupe $Z_{cut\text{-}ab}$ est donné par l'équation suivante:

$$Z_{cut-ab} = \frac{(Z_a + Z_b)}{2} \tag{3.32}$$

Revenons à l'exemple de la figure 3.3. La classification d'un nouvel individu P dont le score de discrimination est Z_p est défini comme suit:

$$\begin{cases} si \quad Z_p > \mathrm{Z}_{\text{cut-ab}} \quad alors \quad P \in B \\ \text{sinon} \qquad\qquad\qquad\quad P \in A \end{cases} \tag{3.33}$$

Dans le cas où si les classes ne possèdent pas le même nombre de données d'apprentissages, le score de la coupe défini dans l'équation ne peut séparer efficacement les deux classes. Nous associons alors un poids à la moyenne de chaque classe et nous redéfinissons l'équation permettant le choix optimal du score de coupe $Z_{cut\text{-}ab}$ comme suit:

$$Z_{cut-ab} = \frac{(n_a.Z_a + n_b.Z_b)}{n_a + n_b} \tag{3.34}$$

n_a et n_b représentent le nombre de données d'apprentissage appartenant respectivement aux classes A et B [121].

Il est bien possible d'étendre l'approche de la régression linéaire discriminante de telle manière à effectuer la séparation entre plusieurs classes en définissant une coupe discriminante entre chaque couple de classes adjacentes. Cette méthode a été utilisée par Xiaoguang Lu and Anil K. Jain [142] dans le cadre de l'identification de personnes et la classification ethnique à partir d'images faciales en couleur.

3.8 La régression logistique

3.8.1 Introduction

La régression est une méthode incontournable en traitement des données en particulier dans une démarche de modélisation. Elle consiste à mettre en relation une variable à expliquer Y avec une ou plusieurs variables explicatives appelées prédicteurs. La méthode la plus utilisée est la régression linéaire qui s'applique lorsque la variable à expliquer est, comme les variables explicatives, continue [69].

Lorsque la variable Y n'est pas continue mais traduit l'appartenance à un groupe, il devient incorrect d'employer la régression classique à des fins de modélisation et de prévision [103]. Un cas particulier est celui pour lequel la variable Y ne prend que deux valeurs qui signifient l'appartenance à une catégorie ou à un groupe d'individus. Dans ce cas, la méthode statistique adaptée est la régression logistique binaire. Il est également concevable d'utiliser l'analyse discriminante linéaire dont l'objet essentiel est le classement c'est-à-dire l'affectation d'individus à des groupes prédéfinis. Cependant, lorsque les conditions d'application de l'analyse discriminante ne sont pas réunies, il est préférable d'employer la régression logistique binaire [72] [167] [202] qui est sans doute la forme la plus courante de régression logistique [176].

Le développement de modèles pour données binaires a été favorisé par les besoins des biologistes [49]. En 1936, Fisher employait les termes d'analyse discriminante [106]. Le domaine médical s'est ensuite intéressé à ces techniques pour des études épidémiologiques, des pronostics de maladie, etc. [49]. Cox a joué un rôle fondamental dans la formulation du modèle de régression logistique, comme en témoigne, par exemple son ouvrage écrit en 1970 [55] ou celui écrit conjointement avec Snell [56]. Actuellement, c'est dans les domaines médical et pharmaceutique que la régression logistique est la plus utilisée. L'emploie de la régression logistique est aussi présente dans d'autres domaines comme en économie, en agronomie et en sociologie [69].

3.8.2 Modèle logistique

La régression logistique linéaire (*LLR*) est une méthode standard en classification supervisée, suite au développement des *modèles linéaires généralisés* [150][94][6]. Elle est la méthode de type discriminatif de référence. l'objectif est de prédire et/ou expliquer une variable catégorielle *Y* à partir d'une collection de descripteurs $X = (X_1, X_2, \ldots, X_J)$. Il s'agit en quelque sorte de mettre en évidence l'existence d'une liaison fonctionnelle sous-jacente (en Anglais, *underlying concept*) de la forme:

$Y = f(X,\alpha)$ entre ces variables.

La fonction *f* est le *modèle de prédiction*, on parle aussi de classifieur; α est le vecteur des paramètres de la fonction, on doit en estimer les valeurs à partir des données disponibles [188].

Dans le cadre de la discrimination binaire, nous considérons que la variable dépendante *Y* ne prend que 2 modalités: la valeur "1" ou la valeur "0" (variable binaire). Chacune de ces valeurs correspond à une classe prédéfinie. Nous cherchons à prédire correctement les valeurs de *Y*, mais nous pouvons également vouloir quantifier la probabilité d'un individu à appartenir à la classe 1 ou 0.

3.8.2.1 Un cadre bayésien pour l'apprentissage supervisé

Le classifieur bayésien est celui qui répond de manière optimale aux spécifications ci-dessus. Pour un individu *ω*, il s'agit de calculer les probabilités conditionnelles (probabilité à posteriori) $P[Y(\omega)=y_k / X(\omega)]$ pour chaque modalité y. On affecte à l'individu la modalité la plus probable y_k* c'est à dire $y_k* = \arg\max_k P[Y(\omega) = y_k / X(\omega)]$. On associe donc l'individu à la classe la plus probable compte tenu de ses caractéristiques X(ω).

Pour rendre calculable la quantité $P(Y = y_k /X)$, il nous faudra donc introduire une ou plusieurs hypothèses sur les distributions. En effet, lors du traitement d'un problème réel, il faudrait en toute rigueur s'assurer de la crédibilité des hypothèses avant de pouvoir mettre en œuvre la technique [170].

Reconsidérons la probabilité conditionnelle $P(Y = y_k/X)$:

$$P(Y = y_k / X) = \frac{P(Y = y_k) \times P(X / Y = y_k)}{P(X)} = \frac{P(Y = y_k) \times P(X / Y = y_k)}{\sum_k P(Y = y_k) \times P(X / Y = y_k)} \quad (3.35)$$

Dans le cas à deux classes, nous devons comparer simplement *P*(*Y = 1/X*) et *P*(*Y = 0/X*). Formons en le rapport:

$$\frac{P(Y = 1/X)}{P(Y = 0/X)} = \frac{P(Y = 1)}{P(Y = 0)} \times \frac{P(X / Y = 1)}{P(X / Y = 0)} \quad (3.36)$$

La règle de décision devient:

$$\text{Si } \frac{P(Y = 1/X)}{P(Y = 0/X)} > 1 \quad \text{alors} \quad Y=1 \quad (3.37)$$

Le rapport $\frac{P(Y = 1)}{P(Y = 0)}$ est facile à estimer dès lors que l'échantillon est issu d'un tirage aléatoire dans la population, indépendamment des classes d'appartenance des individus. Il suffit de prendre le rapport entre le nombre d'observations de la classe 1 et la classe 0 : $\frac{n_1}{n_0}$.

Le véritable enjeu réside donc dans l'estimation du rapport de probabilité $\frac{P(X / Y = 1)}{P(X / Y = 0)}$ [170]. La régression logistique introduit l'hypothèse fondamentale suivante:

$$\log\ [\frac{P(X / Y = 1)}{P(X / Y = 0)}] = b_0 + b_1 X_1 + ... + b_j X_j \quad (3.38)$$

Cette hypothèse couvre une large palette de lois de distribution des données [14] (page 64) comme la loi normale (comme pour l'analyse discriminante), les lois exponentielles, les lois discrètes, les lois Beta, les lois Gamma et les lois de Poisson.

3.8.2.2 Le modèle LOGIT

La régression logistique peut être décrite d'une autre manière selon un modèle appelé modèle *logit*. Pour un individu *ω*, on appelle transformation *LOGIT* de *π*(*ω*) l'expression ([104] (page 6)):

$$\log\ [\frac{\pi(\omega)}{1 - \pi(\omega)}] = a_0 + a_1 X_1 + ... + a_j X_j \quad (3.39)$$

La quantité $\frac{\pi}{1-\pi} = \frac{P(Y=1/X)}{P(Y=0/X)}$ exprime un *odds* c'est-à-dire un rapport de chance. Par exemple, si un individu présente un *odds* de 2, cela veut dire qu'il a 2 fois plus de chances d'appartenir à la classe 1 qu'à la classe 0.

Posons $C(X) = a_0+a_1X_1+\ldots. +a_jX_j$, nous pouvons revenir sur π avec la fonction logistique:

$$\pi = \frac{e^{C(X)}}{1+e^{C(X)}} = \frac{1}{1+e^{-C(X)}} \quad (3.40)$$

La fonction LOGIT= C(X) est théoriquement défini entre -∞ et +∞. En revanche, $0 \leq \pi \leq 1$ issue de la transformation de $C(X)$ représente une probabilité avec les propriétés inhérentes à une probabilité, entre autres $P(Y = 1/X) + P(Y = 0/X) = 1$.

La règle d'affectation peut être basée sur π de différentes manières. Voici quelques exemples de règle d'affectation:

Si $\frac{\pi}{1-\pi} > 1$ alors Y=1 (3.41)

Si $\pi > 0.5$ alors Y=1 (3.42)

Elle peut être aussi basée simplement sur $C(X)$ avec:

Si $C(X) > 0$ alors $Y = 1$ (3.43)

$C(X)$ et π permettent tous deux de "scorer" les individus, et par là de les classer selon leur propension à être "positif". Cette fonctionnalité est très utilisée dans le ciblage marketing par exemple. On parle de "scoring".

La fonction de transfert logistique est non linéaire (Figure 3.4.a), c'est en ce sens que l'on qualifie la régression logistique de régression non-linéaire dans la littérature.

La figure suivante illustre le tracé de la probabilité conditionnelle P(Y/X) en fonction de la fonction *LOGIT* notée sur le schéma f(x) (tracé de la fonction logistique) et une présentation du résultat de la régression logistique qui construit une séparation linéaire (en violet) dans l'espace de présentation [200].

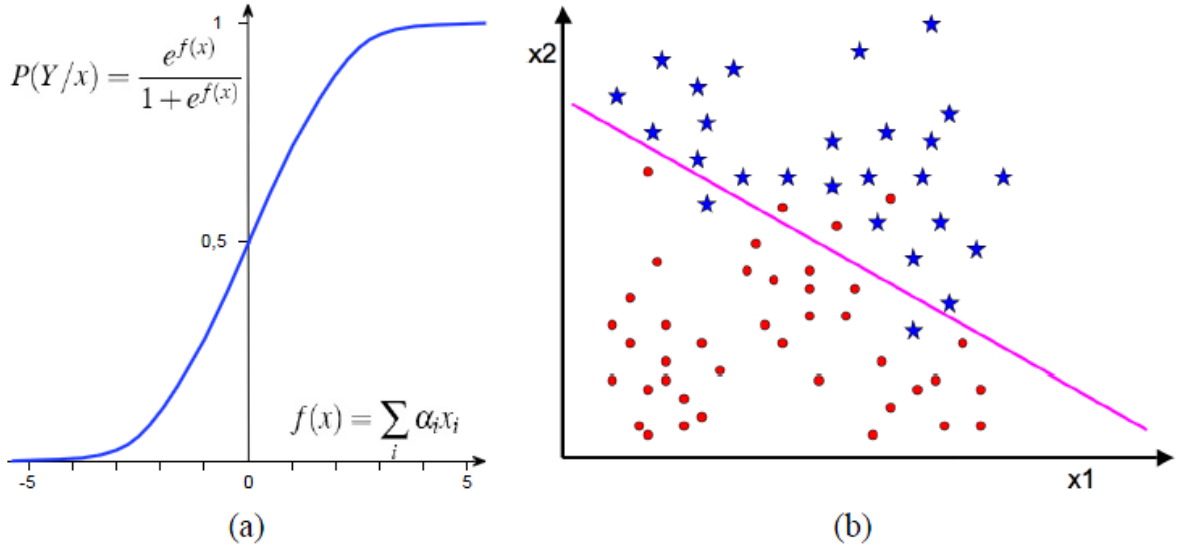


Figure. 3.4: (a) Tracé de la fonction logistique. (b) Illustration 2D du résultat de la régression logistique qui construit une séparation linéaire (en violet) dans l'espace de présentation [200].

3.8.2.3 Equivalence entre les approches

Les deux approches ci-dessus correspondent à deux facettes d'un même problème. En effet:

$$\log[\frac{\pi}{1-\pi}] = a_0 + a_1 X_1 + \ldots + a_j X_j$$
$$= \log[\frac{P(Y=1)}{P(Y=0)} \times \frac{P(X/Y=1)}{P(X/Y=0)}]$$
$$= \log([\frac{P(Y=1)}{P(Y=0)}]) + \log([\frac{P(X/Y=1)}{P(X/Y=0)}])$$
$$= \log\left[\frac{P}{1-P}\right] + (b_0 + b_1 X_1 + \ldots + b_j X_j) \quad (3.44)$$

Les deux formulations (Équations 3.38 et 3.39) sont identiques à une constante près:

$$a_0 = \log\left[\frac{P}{1-P}\right] + b_0 \quad (3.45)$$

Le modèle le plus utilisé est celui du *logit* surtout quand les distributions des variables dépendantes X_i sont inconnues. Dans ce cas, on calcule directement les probabilités conditionnelles. On parle alors d'un apprentissage discriminatif.

Contrairement à la régression linéaire discriminante, dans ce cas nous n'avons pas besoin de connaitre la distribution des individus dans les classes considérées [170].

3.8.2.4 Estimation des paramètres par la maximisation de la vraisemblance

Pour estimer les paramètres de la régression logistique par la méthode du maximum de vraisemblance, nous devons tout d'abord déterminer la loi de distribution de $P(Y/X)$. Y est une variable binaire définie dans $\{1, 0\}$. Pour un individu ω, on modélise la probabilité à l'aide de la loi binomiale $B(1, y)$, avec

$$P[Y(\omega) / X(\omega)] = \pi(\omega)^{y(\omega)} . (1-\pi(\omega))^{(1-y(\omega))} \quad (3.46)$$

Cette modélisation est cohérente avec ce qui a été dit précédemment, en effet:

$$\text{Si } y(\omega) = 1, \text{ alors } P[Y(\omega) = 1/X(\omega)] = \pi \quad (3.47)$$

$$\text{Si } y(\omega) = 0, \text{ alors } P[Y(\omega) = 0/X(\omega)] = 1 - \pi \quad (3.48)$$

3.8.2.4.1 Vraisemblance

La vraisemblance (en Anglais *likelihood*) d'un échantillon Ω s'écrit:

$$L = \prod_{\omega} \pi(\omega)^{y(\omega)} \times (1 - \pi(\omega))^{(1-y(\omega))} \quad (3.49)$$

Pour alléger l'écriture, nous utiliserons pour la suite:

$$L = \prod_{\omega} \pi^{y} \times (1 - \pi)^{(1-y)} \quad (3.50)$$

La vraisemblance correspond à la probabilité d'obtenir l'échantillon Ω à partir d'un tirage dans la population. Elle varie donc entre 0 et 1. La méthode du maximum de vraisemblance consiste à produire les paramètres $a = (a_0, a_1, \ldots, a_J)$ de la régression logistique qui rendent maximum la probabilité d'observer cet échantillon [156] (page 81).

3.8.2.4.2 Log-vraisemblance

Pour faciliter les manipulations, on préfère souvent travailler sur la log-vraisemblance notée *LL* (log-likelihood) qui est définie comme suit:

$$LL = \sum_{\omega} y \times \log \ \pi + (1 - y) \times \log \ (1 - \pi) \quad (3.51)$$

Le logarithme étant une fonction monotone, le vecteur a qui maximise la vraisemblance est le même que celui qui maximise la log-vraisemblance. Cette dernière en revanche varie entre −∞ et 0 [170].

3.8.2.4.3 Déviance

Comme alternative à la log-vraisemblance, bien souvent, on effectue la minimisation de la quantité:

$D_M = -2 \,.\, LL$ (3.52)

appelée *déviance* [172] (page 13). Contrairement à la log-vraisemblance, elle est positive. L'objectif de l'algorithme d'optimisation est de minimiser cette déviance.

Dans certains ouvrages, on définit la déviance *D* de manière plus générique ([172], page 13; [32], page 405; [137], page 115):

$$D = 2 \times \log\left[\frac{L(Modèle\ saturé)}{L(Modèle\ étudié)}\right]$$

$$= -2 \times LL(Modèle\ étudié) - \left[-2 \times LL(Modèle\ saturé)\right]$$

$$= \mathrm{D_M} - \left[-2 \times LL(Modèle\ saturé)\right]$$

$$= -2\sum_{\omega}\left[\mathrm{y}\log\left(\frac{\hat{\pi}}{y}\right) + (1-y)\log\left(\frac{1-\hat{\pi}}{1-y}\right)\right] \quad (3.53)$$

Un modèle saturé pour des données individuelles est un modèle reconstituant parfaitement les valeurs des variables indépendantes, c'est-à-dire: $\hat{\pi}(\omega) = y(\omega)$. Sa vraisemblance est égale à 1 (Équation 3.50) et sa log-vraisemblance 0 (Équation 3.51). Dans ce contexte, *D* = *DM* [170].
Plusieurs techniques d'optimisation existent. Les logiciels s'appuient souvent sur l'algorithme de Newton-Raphson [198] (pages 398 à 400) ou de ses variantes (ex. Fisher Scoring) pour cette tâche.

3.8.3 Méthode de Newton-Raphson

L'algorithme de Newton-Raphson est une des méthodes numériques les plus utilisées pour optimiser la log-vraisemblance [198] (page 398), [104] (page 33), [156] (page 162). Il démarre avec une initialisation quelconque du vecteur de paramètre *a*; pour passer de l'étape (*i*) à l'étape (*i* + 1), il se rapproche de la solution finale $\hat{a}$ en utilisant la formule suivante:

$$a^{i+1} = a^{i} - \left(\frac{\partial^2 LL}{\partial a \ \partial a'} \right)^{-1} \times \frac{\partial LL}{\partial a} \tag{3.54}$$

Plusieurs règles d'arrêt sont possibles pour stopper le processus de recherche [170]:

- On fixe à l'avance le nombre maximum d'itérations pour limiter le temps de calcul. C'est souvent bien utile pour éviter les boucles infinies faute de convergence.

- On stoppe les itérations lorsque l'évolution de la log-vraisemblance d'une étape à l'autre n'est pas significative. Pour cela, on fixe souvent une valeur de seuil ϵ. On arrête le processus si l'écart d'une étape à l'autre est plus petit que le seuil.

- On stoppe les itérations lorsque l'écart entre les vecteurs solutions $\hat{a}$ est faible d'une étape à l'autre. Ici également, souvent il s'agit de fixer un seuil à l'avance auquel on compare la somme des écarts aux carrés ou la somme des écarts absolus entre les composantes des vecteurs solutions.

3.8.3.1 Vecteur des dérivées partielles premières de la log-vraisemblance

Dans l'équation 3.54, le vecteur de dimension (*J*+1×1) correspondant aux dérivées partielles premières de la log-vraisemblance $\Delta(a) = \frac{\partial LL}{\partial a}$ est appelé *vecteur score* ou *vecteur gradient*. Voyons-en le détail pour la variable X_j :

$$\Delta\left(a_j\right) = \sum_{\omega} \left[y(\omega) - \pi(\omega) \right] \times x_j(\omega) \tag{3.55}$$

Lorsque la solution a été trouvée c'est à dire le vecteur $\hat{a}$ permettant d'optimiser LL est obtenu, toutes les composantes du vecteur gradient sont égales à 0. La solution annule la dérivée première par rapport aux paramètres [170].

3.8.3.2 Matrice des dérivées partielles secondes de la log-vraisemblance

La matrice des dérivées partielles seconde $H(a) = \left(\frac{\partial LL^2}{\partial a \ \partial a'} \right)$ est appelée *matrice hessienne*. Elle est très importante car son inverse correspond à la matrice des variances covariances des coefficients. *H*(*a*) est de dimension *(J + 1 x J + 1)* d'expression générale:

$$H(j_1, j_2) = \sum_{\omega} x_{j1}(\omega) \times x_{j2}(\omega) \times \pi(\omega) \times \left[1 - \pi(\omega) \right] \tag{3.56}$$

Il est parfois plus commode de passer par une notation matricielle, nous pouvons écrire:

H(a)= -X'VX (3.57)

où *V* est une matrice diagonale de taille (*n* x *n*) composée de $\pi(\omega)$ x (1 − $\pi(\omega)$) [170]. Nous présenterons dans la section suivante l'algorithme d'approximation correspondant à la méthode de Newton-Raphson.

3.8.3.3 Algorithme de Newton Raphson

Considérons dans cette section les notations suivantes:

Y: variable binaire (0 ou 1). Chacune de ces valeur correspond à une classe prédéfinie.

Y': probabilité d'appartenant à la classe notée 1.

1-Y': probabilité d'appartenant à la classe notée 0.

Selon le principe de la régression logistique, nous pouvons écrire:

$$W = \log(Y'/ 1-Y') = b_0+b_1X_1+b_2X_2+.....+b_pX_p \quad (3.58)$$

Où W la fonction logistique, X_i les variables indépendantes et b_i les paramètres de la régression. L'idée est d'utiliser une transformation symétrique de cette probabilité qui associe la valeur 1 à l'infini, 0 à moins l'infini et 0.5 à 0. La transformation logistique effectue précisément cela. Si Y'=1, W=infini; si Y'=0, W=-infini; si Y'=0.5, W=0 et on a W(Y')= -W'(1-Y) (symétrie).

A partir de (3.58), nous avons l'équation suivante

$$Y' = \frac{\exp(W')}{1+\exp(W')} \quad (3.59)$$

Afin de calculer Y', il convient d'abord d'ajuster les paramètres b_i. L'estimation de b (b_0, b_1….b_p)^T va se faire de manière itérative par la méthode de vraisemblance maximale de type log(L) comme expliqué dans la section précédente.
On s'intéresse dans ce cas à l'observation de plusieurs individus. Soit C la matrice n x p qui représente les coordonnées de tout les n individus à traiter sachant que chaque coordonnée est représentée par une variable indépendante.

Soit H la matrice px1 qui représente la différence entre la variable binaire Y et la probabilité d'appartenance à la classe 1 Y' pour chacun des p individus à traiter c'est-à-dire Y-Y'.

Le nombre maximum d'itérations est fixé à l'avance pour limiter le temps de calcul afin d'éviter les boucles infinies faute de convergence. Les itérations sont stoppées lorsque l'écart entre les vecteurs solutions b^i est faible d'une étape à l'autre [170].
Afin d'estimer b= (b_0, b_1, b_2)^T, il convient d'adopter l'algorithme de Newton-Raphson suivant [170]:

1. Spécifier b^0 quelconque, poser $b=b^0$

2. Calculer W'=X. b pour chaque individu à traiter

3. Calculer $Y' = \frac{\exp(W')}{1+\exp(W')}$ puis Y-Y' pour chaque pixel et constituer ainsi le vecteur H px1, p étant le nombre d'individus à traiter durant l'apprentissage.

4. Calculer la matrice de variance: V= diag(Y'(1-Y')). V est diagonale de taille p x p. Chaque valeur de la diagonale concerne un individu précis parmi les p individus d'apprentissage.

5. Calculer les coefficients mis à jour: $b^{k+1}= b^k - (C^TVC)^{-1} C^T H$ (C^T étant la transposée de C)

6. Poser $b=b^{k+1}$, aller à 2 (jusqu'à convergence).

7. La matrice de variance-covariance des coefficients de la régression est $V(b)= (C^TVC)^{-1}$.

3.8.4 Régression logistique multinomiale

3.8.4.1 Approche multinomiale

Lorsque la variable dépendante prend *K* (*K* > 2) modalités, nous sommes dans le cadre de la régression logistique multinomiale. On peut la voir comme une forme de généralisation de la régression logistique binaire.

L'objectif est de modéliser la probabilité d'appartenance d'un individu à une modalité y_k. Nous écrivons:

$$\pi_k(\omega)= P[Y(\omega) = y_k /X(\omega)] \qquad (3.60)$$

avec la contrainte $\sum_k \pi_k(\omega) = 1$

On s'appuie sur la loi multinomiale pour écrire la vraisemblance

$$L = \prod_{\omega} [\pi_1(\omega)]^{y_1(\omega)} \times .... \times [\pi_k(\omega)]^{y_k(\omega)} \qquad (3.61)$$

Où :

$$y_k(\omega) = \begin{cases} 1 & si \quad Y(\omega) = y_k \\ 0 & \text{sinon} \end{cases} \qquad (3.62)$$

Il s'agit bien d'une généralisation. En effet, nous retombons sur la loi binomiale si *Y* est binaire. L'idée de la régression logistique multinomiale est de modéliser (*K*−1) rapports de probabilités (*odds*). Nous prenons une modalité comme référence (en anglais, *baseline outcome*), et nous exprimons les logits par rapport à cette référence (*baseline category logits*) [5] (pages 307 à 317) [104] (pages 260 à 287).
La catégorie de référence s'impose souvent naturellement au regard des données analysées: les non-malades vs. les différents types de maladies; le produit phare du marché vs les produits outsiders; etc. Si ce n'est pas le cas et si toutes les modalités sont sur un pied d'égalité, nous pouvons choisir n'importe laquelle. Cela n'a aucune incidence sur les calculs, seule l'interprétation des coefficients est différente. Par convention, nous décidons que la dernière catégorie Y_K sera la modalité de référence dans cette partie [170]. Le logit pour la modalité y_k s'écrit:

$$C_k = \log \frac{\pi_k}{\pi_K} = a_{0k} + a_{1k} X_1 + ... + a_{jk} X_j \tag{3.63}$$

Nous en déduisons les (K-1) probabilités à posteriori:

$$\pi_k = \frac{e^{C_k}}{1 + \sum_{k=1}^{K-1} e^{C_k}} \tag{3.64}$$

La dernière probabilité π_K peut être obtenue directement ou par différenciation

$$\pi_K = \frac{1}{1 + \sum_{k=1}^{K-1} e^{C_k}} = 1 - \sum_{k=1}^{K-1} \pi_k \tag{3.65}$$

Pour un individu ω, les probabilités doivent vérifier la relation:

$$\sum_{k=1}^{K} \pi_k(\omega) = 1 \tag{3.66}$$

La règle d'affectation est conforme au schéma bayésien:

$$Y(\omega) = y_k^* \quad \Leftrightarrow \arg\max_k \pi_k(\omega) \tag{3.67}$$

3.8.4.2 Estimation des paramètres

Pour estimer les (*K* − 1) x (*J* + 1) coefficients, nous devons optimiser la log-vraisemblance

$$LL = \sum_{\omega} y_1(\omega) \ln \pi_1(\omega) + ...+ y_K(\omega) \ln \pi_K(\omega) \tag{3.68}$$

via l'algorithme de Newton-Raphson. Pour ce faire, nous avons besoin des expressions du vecteur gradient et de la matrice hessienne [170].

Le *vecteur gradient G* est de dimension (*K* − 1) $*$ (*J* + 1) x 1

$$G = \begin{pmatrix} G_1 \\ \vdots \\ G_{K-1} \end{pmatrix} \tag{3.69}$$

Où G_k, relatif à la modalité y_k, est un vecteur de dimension (*J+1*) x1 dont la composante n^o j s'écrit :

$$g_{k,j} = \sum_{\omega} x_j(\omega) \times \left[y_k(\omega) - \pi_k(\omega) \right] \tag{3.70}$$

Concernant la matrice hessienne, elle est de dimension *(J + 1)*(J+1)*(K- 1)*(K- 1).*

$$H = \begin{pmatrix} H_{1,1} & \cdots & H_{1,K-1} \\ \vdots & & \vdots \\ H_{K-1,1} & \cdots & H_{K-1,K-1} \end{pmatrix} \tag{3.71}$$

$H_{i,j}$ est de dimension (*J* + 1) x (*J* + 1), définie par

$$H_{i,j} = X'(w).V.X(\omega) \tag{3.72}$$

où *V* est une matrice diagonale de taille (*n* x *n*) composée de $\pi_i(\omega)$ x (δ_{ij} (ω)– $\pi_j(\omega)$) avec
X(ω) = (1, X_1 (ω), … , X_J (ω)) est le vecteur de description de l'observation ω et δ_{ij} (ω) est défini comme suit:

$$\delta_{ij}(\omega) = \begin{cases} 1 & si \quad i = j \\ 0 & si \quad i \neq j \end{cases} \tag{3.73}$$

La matrice *H* est symétrique c'est à dire $H_{i,j} = H_{j,i}$.

3.9 Régression logistique bayésienne

Le formalisme bayésien est bien adapté pour ajuster les valeurs à priori inconnues des variables ou paramètres utilisés dans un modèle donné. Ce modèle a été proposé par Taeryon Choi et al. en 2006. L'inférence bayésienne pour des analyses logistiques suit le schéma habituel pour toutes les analyses bayésiennes [47]:

1. On formalise d'abord la fonction de vraisemblance des données traitées.

2. Former la distribution à priori de tous les paramètres à priori inconnus.

3. Utiliser le théorème de Bayes pour trouver la distribution à posteriori pour tous les paramètres traités.

3.9.1 La fonction de vraisemblance

La Fonction de vraisemblance est définie dans l'équation 3.50 dans la section 3.7.2.4.1. Cette fonction peut donc être écrite de la manière suivante [47]:

$$L=\prod_{i=1}^{n}\left[\left(\frac{e^{\beta_0+\beta_1X_{i1}+\ldots+\beta_pX_{ip}}}{1+e^{\beta_0+\beta_1X_{i1}+\ldots+\beta_pX_{ip}}}\right)^{yi}\left(1-\frac{e^{\beta_0+\beta_1X_{i1}+\ldots+\beta_pX_{ip}}}{1+e^{\beta_0+\beta_1X_{i1}+\ldots+\beta_pX_{ip}}}\right)^{(1-yi)}\right] \tag{3.74}$$

y_i représente la variable binaire (1 ou 0) qui correspond à la classification de la donnée i.
n représente le nombre de données traitées.
X_{ij} correspond à la $j^{ième}$ variable indépendante de la donnée i.
β_i correspond au $i^{ième}$ paramètre du modèle logistique.
La log vraisemblance se présente donc comme suit:

$$LL=\sum_{i=1}^{n}\left[y_i\log\left(\frac{e^{\beta_0+\beta_1X_{i1}+\ldots+\beta_pX_{ip}}}{1+e^{\beta_0+\beta_1X_{i1}+\ldots+\beta_pX_{ip}}}\right)+(1-y_i)\log\left(1-\frac{e^{\beta_0+\beta_1X_{i1}+\ldots+\beta_pX_{ip}}}{1+e^{\beta_0+\beta_1X_{i1}+\ldots+\beta_pX_{ip}}}\right)\right] \tag{3.75}$$

3.9.2 Distribution à priori

Généralement, n'importe quelle distribution à priori peut être employée, selon l'information disponible. La distribution à priori la plus utilisée pour les paramètres de la régression logistique est celle d'une loi normale et se présente sous la forme suivante:

$$\beta_j \sim \mathrm{N}(\mu_j,\sigma_j^2) \tag{3.76}$$

Le choix le plus commun pour le μ est zéro et σ est souvent choisi dans l'intervalle [10 100] [47].

3.9.3 Distribution à postériori via le théorème de Bayes

La distribution à postériori est obtenue en multipliant la distribution à priori à travers tous les paramètres par la fonction de log vraisemblance. L'équation 3.77 montre un exemple avec une distribution à priori d'une loi normale $\mathrm{N}(\mu_\mathrm{j}, \sigma_\mathrm{j}^2)$ pour les paramètres β.

$$P = \sum_{i=1}^{n} \left[y_i \log\left(\frac{e^{\beta_0+\beta_1 X_{i1}+\ldots+\beta_p X_{ip}}}{1+e^{\beta_0+\beta_1 X_{i1}+\ldots+\beta_p X_{ip}}} \right) + (1-y_i) \log\left(1 - \frac{e^{\beta_0+\beta_1 X_{i1}+\ldots+\beta_p X_{ip}}}{1+e^{\beta_0+\beta_1 X_{i1}+\ldots+\beta_p X_{ip}}} \right) \right]$$

$$\times \prod_{j=0}^{p} \frac{1}{\sqrt{2\pi}\sigma_j} \quad \exp\left\{ -\frac{1}{2}\left(\frac{\beta_j - \mu_j}{\sigma_j} \right)^2 \right\} \qquad (3.77)$$

L'ajustement optimal des paramètres β se fait par minimisation de la distribution à posteriori en adoptant des algorithmes destinés à cette tâche. L'algorithme le plus utilisé est celui de Newton Raphson présenté dans la section 3.7.3.3.

Riadh Ksantini et al. ont eu recours à ce type de méthode dans [131] en 2008 dans le domaine de recherche d'images en couleur et ont constaté la supériorité de cette approche sur beaucoup de méthodes de classification linéaire dont la régression logistique.

3.10 Régression logistique à noyaux

3.10.1 Astuce du noyau (Kernel Trick)

Depuis de nombreuses années les méthodes à noyau ont démontré leur fort potentiel en discrimination ou régression, comme par exemple le classifieur *Support Vector Machine* [208], *kernel PCA* [181] ou bien encore *kernel Fisher discriminant* [151]. L'intérêt de la plupart de ces approches réside dans la définition de la fonction de prédiction à l'aide d'une combinaison linéaire pondérée de fonctions à noyau. La méthode Support Vector Machine, en particulier, est applicable sur de nombreux problèmes, avec une grande efficacité [13].

L'astuce du noyau (*kernel trick* en anglais) est à l'origine de l'engouement pour les méthodes à noyau. Elle tire partie de la possibilité de calculer directement des produits scalaires dans des espaces de grandes dimensions, résultant de transformations non linéaires des variables initiales, sans projeter explicitement les points de l'ensemble d'apprentissage dans ces espaces. Grâce à cette astuce, n'importe quel algorithme reposant sur la notion de produit scalaire peut être étendu pour traiter des cas non linéaires, en remplaçant le produit scalaire usuel par un noyau dont la définition est donnée ci-après [52].

Un noyau est une fonction de X x X → R symétrique, telle que:

$K(x_i, x_j) = <\Phi(x_i), \Phi(x_j)>, \quad \forall x_i, x_j \in X$ (3.78)

avec Φ : X → F, où F est un espace muni d'un produit scalaire.

F est l'espace de projection et est connu sous le nom d'espace d'Hilbert.

Un noyau est donc un produit scalaire, c'est-à-dire une forme bi-linéaire symétrique définie positive. Pour manipuler un nuage de points, il n'est pas forcément nécessaire de connaître la position de chacun d'entre eux dans l'espace, il suffit souvent de pouvoir calculer leurs produits scalaires. Le noyau polynomial permet par exemple de travailler dans un espace résultant de transformations polynomiales de degré 2 des variables initiales comme le montre l'exemple suivant:

Soit x = (x_1, x_2) et $\Phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$

On peut définir:

$$\begin{aligned} k(x_i, x_j) &= < x_i, x_j >^2 \\ &= (\mathrm{x_{i1}}.\mathrm{x_{j1}} + \mathrm{x_{i2}}.\mathrm{x_{j2}})^2 \\ &= \mathrm{x_{i1}^2}..x_{j1}^2 + x_{i2}^2.x_{j2}^2 + 2.\ x_{i1}..x_{j1}.\ x_{i2}..x_{j2} \\ &= < \Phi(x_i), \Phi(x_j) > \end{aligned} \quad (3.79)$$

Grâce aux noyaux, il est donc possible de travailler dans des espaces résultants de transformations non linéaires des variables initiales; cela peut avoir un impact important sur la solution trouvée, comme le montre la figure 3.5. Nous pouvons observer sur celle-ci qu'un problème de discrimination qui n'était pas linéaire dans l'espace initiale peut le devenir dans l'espace transformé [52].

La frontière de décision dans l'espace d'origine (a) correspond à une ellipse. La figure de droite présente la projection des mêmes individus dans le plan définis par les deux premières composantes de l'espace noyau générées par le noyau polynomial de degré 2, la frontière de décision est linéaire dans ce plan (b).

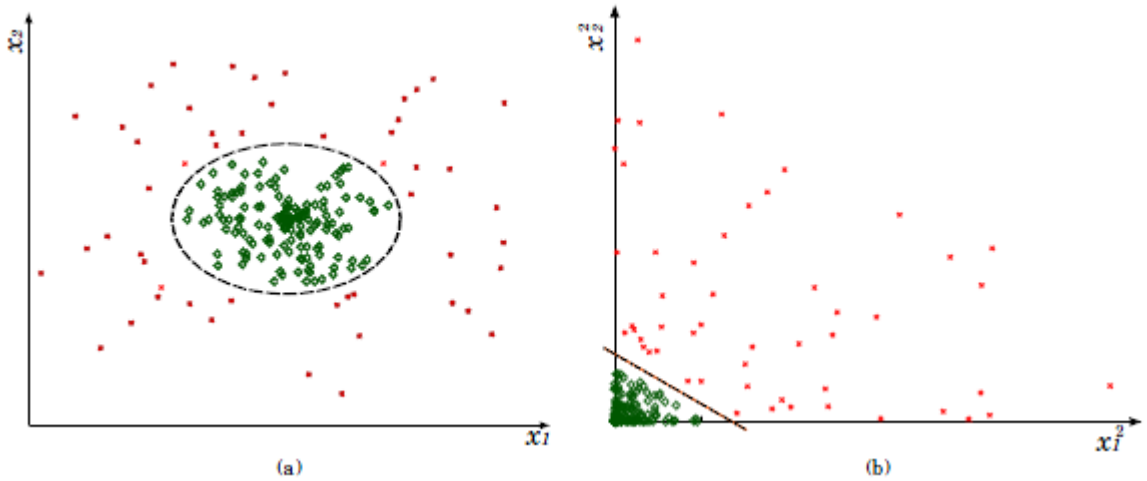


Figure. 3.5: Exemple de changement de représentation transformant une frontière de décision non linéaire en frontière de décision linéaire [52].

Dans le cadre des méthodes à noyau, les données ne sont plus manipulées à l'aide de la matrice d'observation X contenant les différentes variables pour chacun des individus, mais grâce à la matrice contenant les différentes valeurs du noyau pour tous les couples d'individus. Cette matrice, appelée matrice de Gram, est définie formellement de la manière suivante:

Soit un jeu de données $X = \{x_1, x_2, \ldots x_n\}$ et un noyau k, la matrice de Gram G associée, est la matrice telle que:

$$G_{ij}=k(x_i, x_j), \quad \forall i,j \in \{1,\ldots,n\} \tag{3.80}$$

Les algorithmes associés aux méthodes à noyaux, utilisent en général en entrée une matrice de Gram décrivant les données grâce aux valeurs prises par le noyau lorsqu'il est évalué sur les différents exemples de l'ensemble d'apprentissage. A partir de celle-ci, une classification peut être effectuée sur de nouvelles données à traiter. En utilisant cette nouvelle représentation du jeu de données, les méthodes à noyaux dissocient clairement représentation des données et exploitation des données [52].

De nombreuses méthodes linéaires peuvent être reformulées de manière à ne faire intervenir que des produits scalaires entre individus et celles-ci peuvent donc être étendues pour traiter des données projetées dans un espace résultant de transformations non-linéaires des variables initiales grâce à l'astuce du noyau. Les SVM [21] utilisent ainsi l'astuce du noyau pour produire des frontières de décision non-linéaires, de même que la régression logistique [231] qui peut elle aussi être *kernélisée*. Nous allons présenter dans la section suivante la régression logistique à noyaux.

3.10.2 Régression logistique à noyau

La régression logistique à noyau est un modèle qui repose à la fois sur l'utilisation de la régression logistique et les méthodes à noyau. Sans passage par une sélection d'une partie des attributs des données (dans la plupart des techniques d'apprentissage) pour réduire la dimension de l'espace d'entrée, ce type de modèle impose que le nombre des paramètres soit linéairement lié au nombre des données d'apprentissage. Contrairement à la régression logistique classique, cette méthode est donc non paramétrique [101].

La régression logistique à noyau est un modèle dont le log-ratio des probabilités conditionnelles est défini comme suit [101]:

$$\log \frac{P(Y=1/X)}{1-P(Y=1/X)} = f(X)+b \tag{3.81}$$

Où f est une fonction appartenant à l'espace de projection à noyau (espace de Hilbert) H.
Y est la variable binaire de sortie (variable endogène) et X correspond aux variables indépendantes de l'individu traité. La fonction *f* est définie comme suit [35]:

$$f(X)=\sum_{i=1}^{n}\alpha_i k(X,X_i) \tag{3.82}$$

Avec $\alpha_i = \frac{1}{\lambda}\left(\frac{y_i+1}{2} - P(Y=1/x_i)\right)$ et k la fonction noyau.

Le modèle log-vraisemblance se présente sous la forme suivante [201]:

$$L=\sum_{i=1}^{n}\log\left(1+e^{-yi(f(xi)+b)}\right)+\frac{\lambda}{2}\alpha^T K\alpha \tag{3.83}$$

où λ est un hyper paramètre, qui peut être ajusté par expérimentation, α le vecteur des paramètres α_i et K représente la matrice de Gram. L'ajustement optimal des paramètres α_i revient à minimiser l'équation 3.83 avec α_i différent de 0.

Une fois les paramètres α_i ajustés d'une manière optimale, la classification finale se fait selon l'équation suivante [101]:

$$P(y_i = 1/X) = \frac{\exp(\sum_{i=1}^{n} \alpha_i k(X, Xi) + b)}{1 + \exp(\sum_{i=1}^{n} \alpha_i k(X, Xi) + b)} \quad (3.84)$$

Avec x l'individu à classifier et x_i le $i^{ème}$ individu de la base d'apprentissage.

Riadh Ksantini et al. [131] ont proposé une méthode d'apprentissage basée à la fois sur les noyaux de Fisher, le théorème de Bayes et la régression logistique. Cette méthode est plus connue sous le nom de *régression linéaire bayésienne à noyaux* (RLBN) (voir l'annexe). Nous nous somme basés sur ce type d'approche pour construire notre modèle.

3.11 Les SVM (Support Vector Machine)

Les SVM (Support Vector Machine) représentent un modèle qui se base aussi sur la méthode des noyaux. Ils sont issus d'une formulation de la théorie de l'apprentissage statistique due en grande partie à l'ouvrage de Vapnik en 1995 intitulé "*the nature of learning statistical theory*" [210]. Les SVM, conçus au départ pour une séparation des données en deux classes, constituent un modèle discriminant qui tente de minimiser les erreurs d'apprentissage tout en maximisant la distance entre les deux classes. Il s'agit en fait de trouver un classifieur linéaire dit hyperplan qui va séparer au mieux les données (voir figure 3.6). Les points les plus proches (points entourés dans la figure 3.6) sont appelés les vecteurs de support. Ces seuls points sont utilisés pour déterminer l'hyperplan [109].

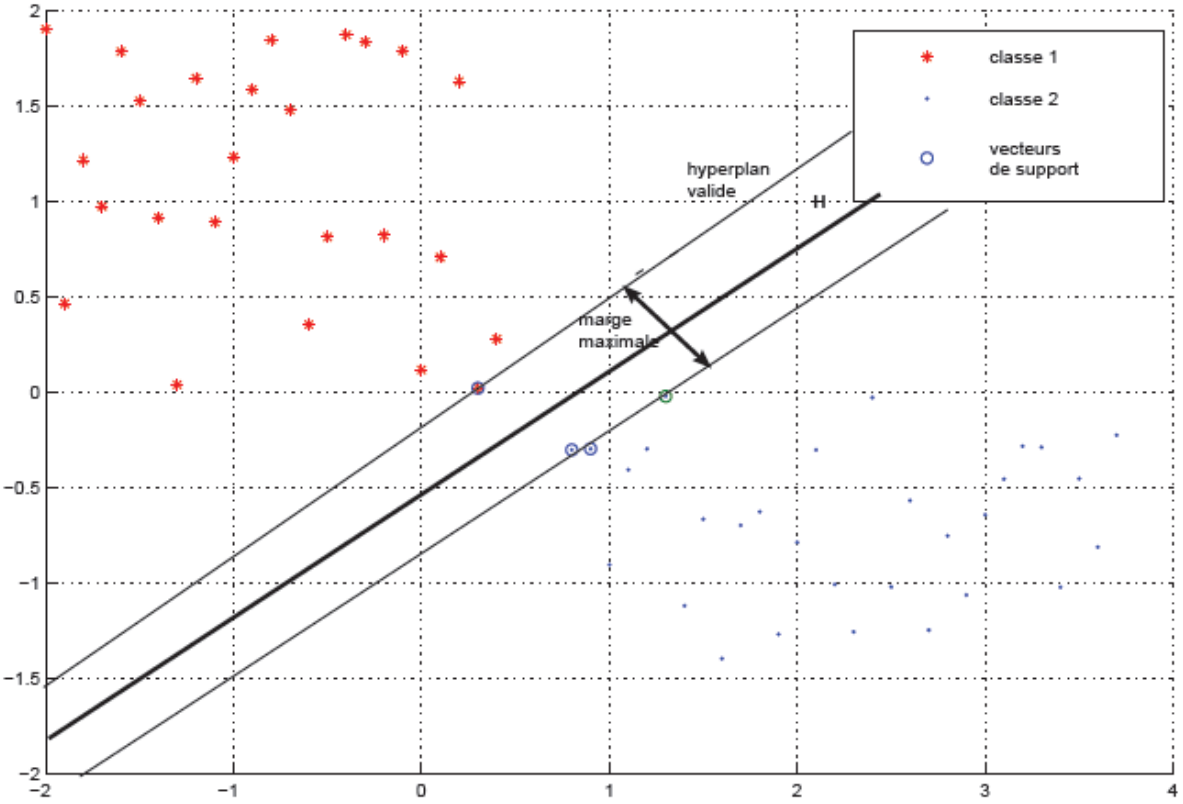


Figure. 3.6: Schéma de l'hyperplan séparant deux classes [109].

Vu la multitude des hyperplans valides, séparant les deux classes, les SVM ont la propriété particulière de chercher l'hyperplan optimal (H dans la figure 3.6). Cet hyperplan correspond à celui qui passe *au milieu* des points des deux classes. Plus formellement, cela revient donc à chercher l'hyperplan dont la distance minimale aux données d'apprentissage est maximale. Autrement dit, celui qui se trouve le plus loin des points. La distance entre l'hyperplan et le point qui lui est le plus proche (vecteur de support) est appelée *marge*. Pourquoi chercher une marge maximale? La réponse est simple: supposons qu'un point n'a pas été décrit parfaitement, une petite variation ne changera pas sa classification si sa distance à l'hyperplan est grande. Dans le cas inverse où la marge est faible, la classification d'un nouvel élément risque d'être erronée. Donc une marge maximale garantit plus de sécurité lorsqu'on veut classer un nouvel élément. De nombreux travaux ont démontré la supériorité des SVM par rapport aux méthodes discriminantes classiques [109]. Dans le cas des SVM, la variable binaire de sortie prend généralement ses valeurs dans l'ensemble {-1, 1}.

La fonction de vraisemblance dans ce cas est définie comme suit:

$$L = (1 - y_i f(x_i)) + \frac{\lambda}{2} \alpha^T K \alpha \qquad (3.85)$$

Avec *f(x)* est la même fonction présentée pour le modèle de régression logistique à noyau notée de la même manière. x_i, i∈ [1, n] représente les données d'apprentissage et y_i la variable de sortie binaire qui prend ses valeurs dans l'ensemble {-1,1}. Chacune de ces valeurs correspond à une classe bien définie. L'ajustement optimal des paramètres du vecteur α revient à minimiser la fonction 3.85. Cette minimisation assure la maximisation de la marge de séparation entre les deux classes considérées. Le modèle de classification d'un nouvel individu à traiter z est défini par l'équation suivante [231]:

$$P(z) = \sum_{i=1}^{n} \alpha_i y_i k(x_i, z) + b \tag{3.86}$$

$$\begin{cases} SiP(z) \geq 1 \;\; alorsY = 1 \\ SiP(z) \leq -1 \;\; alorsY = -1 \end{cases} \tag{3.87}$$

Avec Y représente la classe (-1 ou 1) du nouvel individu à classifier. Daprès Ji zhu et Treveor Hasti [232], ce type de méthode offre des performances de classification similaires à celles de la régression logistique à noyau.

3.12 Conclusion

L'apprentissage est une problématique issue initialement des travaux axés sur l'intelligence artificielle dès la fin des années 70 lorsque la vogue des systèmes experts se heurte à la difficulté d'acquérir l'expertise existante, ou de la constituer lorsqu'elle est inexistante. Depuis cette date, la communauté scientifique a proposé une série de techniques et d'outils pour l'apprentissage basés sur les statistiques. Son objectif vise à construire des hypothèses à partir d'exemples. Ce problème se décline en deux variantes: l'approche supervisée et l'approche non supervisée. Dans la première, on connait les classes possibles et on dispose d'un ensemble d'objets déjà classés, servant d'ensembles d'apprentissage. Le problème est alors d'être capable d'associer à tout nouvel objet sa classe la plus appropriée, en se servant d'objets déjà étiquetés (déterminer, par exemple, à partir d'une base de données exprimant les problèmes cardiovasculaires d'un ensemble de patients, quels risques cardio-vasculaires a un nouveau patient). Dans la seconde par contre (l'apprentissage non supervisée), les classes possibles ne sont pas connues à l'avance, et les exemples disponibles ne sont pas étiquetés. Le but est alors de regrouper dans un même cluster (groupe) les objets présentant une certaine similarité entre eux, pour constituer les classes en étudiant la structure des données d'apprentissage. Nous distinguons dans les problèmes d'apprentissage supervisé, les approches discriminatives et les approches génératives. La modélisation discriminative modélise directement la règle de classification sans pour autant modéliser la structure du système. La modélisation générative cherche en premier lieu à trouver une structure pour la distribution jointe des variables d'entrée et de

sortie. En ce sens, un travail de modélisation supplémentaire est nécessaire par rapport aux approches discriminatives. Nous avons appuyé la théorie sous-jacente à ces méthodes par des exemples explicatifs afin d'affermir la compréhension.

Le chapitre suivant (chapitre 4) présentera le processus complet de la construction de notre modèle de peau basé sur la régression logistique, les noyaux de Fisher et l'approche bayésienne.

CHAPITRE 4
APPROCHE ADOPTEE ET EVALUATION EXPERIMENTALE

4.1 Introduction

La détection de la peau est une étape importante dans de nombreux algorithmes de vision par ordinateur. Il est habituellement un processus qui commence au niveau du pixel, et qui implique un pré-processus de transformation colorimétrique suivie par un processus de classification. Il s'agit d'une tâche préliminaire de base utilisée dans de nombreux domaines comme en télésurveillance et en reconnaissance de forme. Nous présentons dans ce chapitre un détecteur de peau basé sur une approche nommée régression logistique bayésienne à noyau (RLBN) [131]. Dans le cadre de cette approche, le noyau de Fisher est exploité pour faire une projection des données d'origines présentées dans un espace dans lequel il n'est pas toujours possible de séparer les différentes classes linéairement dans un nouvel espace de dimension généralement plus grande de telle manière à pouvoir discriminer les classes (classe peau et classe non peau) linéairement. La régression logistique sert de modèle de prédiction et s'emploie dans le cadre de la discrimination des pixels peau et ceux de non peau. Le théorème de Bayes est utilisée afin d'estimer les valeurs optimales des paramètres de l'équation logistique. Nous nous sommes intéressés à traiter trois types de couleurs de peau appartenant à différentes races ethniques: la race jaune, la race noire et la race blanche. A cet effet, nous avons construit un classifieur spécifique à chaque race puis nous avons combiné les trois classifieurs en un seul dit classifieur de synthèse destiné à la détection de tout type de peau. Nous avons utilisé la base d'image Compaq pour évaluer notre classifieur et les résultats obtenus surpassent ceux obtenus par les classifieurs basés sur des méthodes proposées précédemment et évalués sur la même base d'images que la notre.

Nous présenteront dans la section 4.2 la base d'image utilisée (*Compaq*), puis nous nous intéresserons aux performances des méthodes existantes ayant utilisée cette base d'images dans la section 4.3 avant de nous attarder dans la section 4.4 sur la présentation de notre système pour la détection de tout type de peau. La section 4.5 abordera les évaluations expérimentales effectuées en comparaison des résultats obtenus par les méthodes proposées précédemment et évalués sur la même base d'images que la notre.

4.2 La base d'image Compaq

La construction de la base d'apprentissage est un élément important dans une démarche d'extraction de connaissances à partir des données. Pour un problème de classification de pixels en peau ou non, la qualité du corpus peut être jugée sur les facteurs suivants:

1. la taille de la base d'images et la variété dans les contenus des scènes, qui doit être représentative pour les différents sexes, races, et conditions d'éclairages.

2. la source de la base d'images, qui dépend particulièrement de l'application, tels que le filtrage de sites adultes sur le web et la détection de visage dans la vidéo.

Nous utilisons un corpus large d'images, contrairement à la majorité des travaux existants, qui pratiquent une phase d'apprentissage sur des classes prédéfinies d'images, sous des conditions d'éclairage connues à l'avance [96]. Si ces modèles conviennent généralement à des systèmes à base d'images spécifiques donc peu variées par exemple aux conditions d'éclairage, ils sont peu adaptés aux systèmes contenant une grande variété d'images.

Pour l'évaluation des performances des différents modèles de couleur de peau, des conditions de test doivent rester identiques. Malheureusement, il existe de nombreuses méthodes de détection de peau qui fournissent des résultats en se basant sur leurs propres bases d'images qui ne sont malheureusement pas accessibles. La base d'images la plus célèbre pour l'apprentissage et les tests est celle de Compaq [117].

Nous avons utilisé cette base d'image pour valider notre méthode de détection de la peau. Celle-ci inclut environ 4660 images contenant des pixels peau et 9000 ne contenant pas de pixels peau. Une vérité terrain de 4600 images binaires correspondant aux images contenant des pixels peau est disponible. Celui-ci est utilisé lors de la phase d'apprentissage pour identifier les parties peau de celles de non peau dans les images. Toutes les images de cette base sont obtenues à partir du Web. Différents types de peau appartenant à différentes races à savoir la race blanche, jaune et noire sont présentes sur ces images. La base d'images Compaq est caractérisée par l'hétérogénéité des conditions d'acquisition des images comme l'illumination variante, l'acquisition sous différents angles et positions, la diversité des appareils de capture des images avec différentes caractéristiques ainsi que la diversité des sources d'acquisition des images (capture à travers une vitre colorée ou non, un écran de télévision…). En outre, les personnes représentées sur ces images sont de différents âges (pour plus de détails voir la section 1.2 du premier chapitre). Cette base est donc sujette à toute sorte de distorsion, d'où toute la difficulté de classification des pixels issus des images lui appartenant en pixels peau et non peau. La figure 3.1 montre quelques exemples de ces images.

-Les images (1a) et (1b) sont brouillées.
-L'illumination est très variante dans les images (1c), (1d) et (2b).
-L'image (2d) est acquise à travers une fenêtre
-L'image (2a) est acquise à partir d'un écran de télévision.
-Le pilote sur l'image (2c) porte un casque coloré.

Toutes ces conditions créent des difficultés considérables pour la détection de la peau au sein de la base d'images Compaq.

Figure. 4.1: Quelques images du corpus Compaq.

4.3 Performances des techniques existantes

4.3.1 Types de méthodes existantes

Tous les travaux rapportés sur la détection de la peau ont utilisé la classification comme stratégie de détection. Les techniques de détection basées sur la couleur peuvent être classifiées en trois catégories que nous allons maintenant aborder.

Les méthodes explicites utilisent des règles de décision empiriques et/ou statistiques [212] pour la détection des pixels ayant la couleur de la peau. Les images doivent être acquises dans un environnement contrôlé (i.e. un éclairage réglable) [221]. Une méthode explicite est une méthode de classification qui consiste à définir explicitement les frontières de la région peau (cluster) dans l'espace de couleur utilisé.

Les méthodes non paramétriques visent à estimer la distribution de la couleur de la peau sans une modélisation explicite à partir d'un ensemble d'apprentissage. Elles se traduisent généralement par un histogramme et une carte de probabilité *SPM (Skin Probability Map)* qui affecte une valeur de probabilité à chaque pixel [87], [26] ou des réseaux de neurones ou de cartes *SOM*. Elles ne dépendent pas de la forme de la fonction de distribution de la teinte.

Les méthodes paramétriques estiment la distribution de la couleur de la peau sous forme de modèles explicites. La distribution est généralement caractérisée par une densité ou une somme de densités de probabilité dont les paramètres sont obtenus à partir d'un ensemble d'apprentissage.

4.3.2 Mesures d'évaluation

Dans le cadre de l'évaluation de notre classifieur, nous considérons un problème simple de classification pour lequel nous nous intéressons à identifier l'appartenance de pixels issus de la base d'image (la classe peau ou la classe non peau). Nous indiquons par classe *positive* la classe peau et la classe *négative* la classe non peau. A cet effet, nous pouvons construire la *matrice de contingence* [62] (voir figure 4.1) qui nous indique quatre types d'informations:

1)- le nombre de pixels peau classifiés correctement (nombre de *vrais pixels positifs NVP*), 2)-le nombre de pixels non peau classifiés correctement (nombre de *vrais pixels négatifs NVN*), 3)- le nombre de pixels peau mal classifiés (nombre de *faux pixels négatifs NFN*), 4)- le nombre de pixels non peau mal classifiés (nombre de *faux pixels positifs NFP*).

Cette matrice représente la répartition des pixels pour la tâche de classification binaire (peau et non peau). Les valeurs en gras correspondent aux documents correctement classés.

Tableau. 4.1: Matrice de contingence bi classe.

Classe réelle / Classe trouvée	Classe peau	Classe non peau
Classe peau	**NVP**	NFP
Classe non peau	NFN	**NVN**

A partir de la matrice du tableau 4.1, nous définissons les mesures d'évaluation suivantes:

Le taux de vrais positifs (*VP*) représente la probabilité qu'un pixel appartenant à la classe peau soit affecté à la classe peau. On emploie également souvent le terme de *rappel* pour désigner cette mesure.

Le taux des faux positifs (*FP*) correspond à la probabilité qu'un pixel appartenant à la classe non peau soit affecté à la classe peau.

Les équations dénotant ces mesures se présentent comme suit:

$$VP = \frac{NVP}{NVP + NFN} \tag{4.1}$$

$$FP = \frac{NFP}{NFP + NVN} \tag{4.2}$$

A partir de ces deux taux on peut déduire encore :

Le taux des faux négatifs (FN) qui correspond à la probabilité qu'un pixel appartenant à la classe peau soit affecté à la classe non peau.

Le taux des vrais négatifs (VN) qui correspond à la probabilité qu'un pixel appartenant à la classe non peau soit affecté à la classe non peau.

Les équations dénotant ces mesures se présentent comme suit:

$$FN = \frac{NFN}{NVP + NFN} \tag{4.3}$$

$$VN = \frac{NVN}{NVN + NFP} \tag{4.4}$$

Beaucoup de travaux effectués sur la détection de la peau en utilisant la base d'images Compaq pour la validation des méthodes qui leurs sont consacrées n'utilisent pas les mesures de FN et VN. Pour des fins de comparaison avec ces méthodes, nous utiliserons les mêmes critères d'évaluation et nous omettons donc nous aussi d'utiliser ces deux mesures. Durant l'expérimentation, plus le taux de VP est élevé et le taux de FP est bas, plus la méthode à évaluer est meilleure. Afin de mieux comprendre le fonctionnement de ces deux mesures, nous détaillons sur les tableaux 4.2, 4.3, 4.4 et 4.5 les performances de plusieurs classifieurs *caricaturaux* utilisés sur deux classes peau et non peau de respectivement 100 et 200 pixels.

Tableau. 4.2: Tous les pixels sont de la de la classe peau.

Classe réelle / Classe trouvée	Classe peau	Classe non peau
Classe peau	100	200
Classe non peau	0	0
VP = 100%		
FP = 100%		

Tableau. 4.3: Tous les pixels sont de la classe non peau.

Classe réelle / Classe trouvée	Classe peau	Classe non peau
Classe peau	0	0
Classe non peau	100	200
VP = 0%		
FP = 0%		

Tableau. 4.4: Classifieur *parfait*.

Classe réelle / Classe trouvée	Classe peau	Classe non peau
Classe peau	100	0
Classe non peau	0	200
VP = 100%		
FP = 0%		

Tableau. 4.5: Classifieur *le pire.*

Classe réelle / Classe trouvée	Classe peau	Classe non peau
Classe peau	0	200
Classe non peau	100	0
VP = 0%		
FP = 100%		

Les deux premiers classifieurs représentent des classifieurs *radicaux* qui jugent soit tous les pixels comme étant peau, soit comme étant tous non peau. Les deux autres classifieurs sont des classifieurs *caricaturaux*, soit parfait (tous les pixels sont correctement classés), soit dramatique (les pixels sont toujours mal classés).

Dans [66], les auteurs ont utilisé la base d'image Compaq afin de valider le modèle qu'ils ont proposé. Ils ont utilisé d'autres critères d'évaluation que nous présentons comme suit:

1) *CDR* (*Correct Decision Rate*): pourcentage de pixels correctement classifiés parmi l'ensemble de tous les pixels;

2) *FAR* (*False Acceptance Rate*): pourcentage de pixels non peau mal classifiés parmi l'ensemble de tous les pixels;

3) *FRR* (*False Rejection Rate*): pourcentage de pixels peau mal classifiés parmi l'ensemble de tous les pixels.

A partir de la matrice du tableau 4.1, nous définissons les équations mathématiques dénotant ces mesures comme suit:

$$CDR = \frac{\mathrm{NVP} + \mathrm{NVN}}{\mathrm{NVP} + \mathrm{NVN} + \mathrm{NFP} + \mathrm{NFN}} \tag{4.5}$$

$$FAR = \frac{\mathrm{NFP}}{\mathrm{NVP} + \mathrm{NVN} + \mathrm{NFP} + \mathrm{NFN}} \tag{4.6}$$

$$FRR = \frac{\mathrm{NFN}}{\mathrm{NVP} + \mathrm{NVN} + \mathrm{NFP} + \mathrm{NFN}} \tag{4.7}$$

Plus *CDR* est élevé et *FAR* et *FRR* est petite, plus le système de classification de pixels est meilleur. Les meilleurs résultats obtenus dans [66] sont présentés dans le tableau 4.6.

4.3.3 Analyse comparative des techniques existantes

Le tableau 4.6 illustre les méthodes ayant obtenues les meilleurs résultats sur la base d'images Compaq [117]. Toutes ces méthodes sont destinées à la détection de pixels peau appartenant à n'importe quelle race ethnique. Pour chaque méthode, on recense les références, l'espace de couleur utilisé pour la segmentation, l'intitulé de la méthode, le type de la méthode utilisée ainsi que l'ensemble d'apprentissage et de test utilisés ainsi que deux critères d'évaluation: le *taux de vrai positifs* et le *taux de faux positifs* (hormis la méthode proposée par les auteurs dans [66] qui utilisent les trois derniers critères d'évaluation que nous venons de présenter dans la section précédente). Pour plus de détails sur les types de ces méthodes, voir chapitre 2.

A titre d'exemple, les auteurs Jones et Rehg [117] ont proposé deux méthodes basées sur l'apprentissage pour la détection de la peau; l'un est un modèle Bayésien non paramétrique basé sur les Histogrammes (Bayes SPM), l'autre est paramétrique et se base sur la mixture de gaussiennes. En utilisant la même stratégie d'apprentissage et de test, les auteurs opèrent deux évaluations pour chaque

méthode. Les scores étant assez serrés, nous pouvons noter une légère supériorité de la méthode Bayes SPM.

Tableau. 4.6: Comparaison des différentes méthodes de détection de la peau.

Ref.	Esp. coul	Comp.lumin	Intitulé de la méthode	Type	Ens.Appr/ test	App-renti-ssage	VP	FP
[117]	RGB RGB	Oui Oui	Bayes SPM GMM	Non paramétrique Paramétrique	6822 /6818 images 6210/6818 images	Oui Oui	80 90 90 80	8.5 14.2 15.5 9.5
[27]	TSL	Non	SOM.Map	Non paramétrique	n/a	Oui	78	32
[114]	RGB	Oui	Max.Ent. model	Paramétrique	2 000 000 000 / 2 000 000 000 pixels	Oui	82.9	10
[136]	Xyz YCbCr YIQ	Non Non Non	Ellip.model SGM GMM	Paramétrique Paramétrique Paramétrique	2000/6000 images	Oui Oui Oui	90 90 90	20.9 33.3 30
[26]	RGB YIQ RGB	Oui Oui Oui	Bayes SPM I-axis thresh Thresh.ratio	Non paramétrique Explicite Explicite	n/a	Oui Non Non	93.4 94.7 94.7	19.8 30.2 32.3
[145]	RGB	Oui	Beta mixt-ure models	Paramétrique	2 000 000/ 2 000 000 pixels	Oui	80 90 95	7.2 11.8 19.7
[12]	RGB	Oui	Impl.Math. model	Paramétrique	n/a	Oui	83.3 90.7	15.6 13.3
[4]	RGB	Oui	Color.Dist. Map	Explicite	4000/ 500 images	Non	89.97	9.26
[76]	RGB	Oui	Disc-DT based model	Non paramétrique	4167/9397 images	Oui	90	12.5
[66]	Cb/Cr & Cr	Non	Comb.neu-ral networks	Non paramétrique	420 000/ 420 000 pixels	Oui	CDR 82.61	FAR 14.66	FRR 2.73

Les méthodes explicites proposées par Brand et al. [26] se justifient par la simplicité des règles de détection de la peau, la facilité de leur application et la faible complexité algorithmique. Un bon choix des intervalles de définition des pixels peau mènent à une performance de détection accrue. Cependant, nous remarquons que

deux méthodes explicites sur trois possèdent un taux de faux positifs important. La méthode *distance color map* de M. Abdullah-Al-Wadud et al. [4] apporte une nette amélioration sur le taux de faux positifs qui est beaucoup moins important que dans les modèles de Lee et al. [136]. Selon les scores obtenus, cette méthode surpasse toutes les autres en termes d'efficacité de détection. Le principal inconvénient de ce type de méthode est leur forte dépendance des conditions d'illumination. Les règles de seuillages ne sont efficacement valables que pour des conditions d'illumination précises. Un changement important sur ces conditions implique nécessairement de nouvelles règles à définir. En outre, la complexité à définir ces règles s'accentue si les pixels à traiter correspondent à différentes origines ethniques [46].

Les méthodes non paramétriques ont l'avantage d'être plus stables aux changements dans les conditions d'illumination variantes et la complexité de l'arrière plan et sont connus pour leur rapidité à la fois dans les phases d'apprentissage et de classification [119]. Le modèle Bayésien basé sur les Histogrammes *(Bayes SPM*) donne de meilleurs résultats que les modèles basés sur la distribution gaussiennes dans [117] ainsi que les modèles basés sur la définition de règles explicites dans [26] et démontre une bonne séparabilité entre les pixels peau et non peau quoique nous notons de faibles performances pour les modèles basés sur les réseaux de neurones de type *SOM (Self Organizing Map)*. L'inconvénient majeur de ces modèles est la nécessité d'avoir de larges collections de données pour l'apprentissage afin d'aboutir à de meilleures performances. En outre, ces méthodes sont couteuses en termes d'espace mémoire. La meilleure méthode non paramétrique est celle de El Fkihi S. et al. [76]. Elle se classe parmi les meilleurs travaux avec des performance de détection assez proches que celles de M. Abdullah-Al-Wadud et al. [4] et de Zhanyu Ma et al. [145].

Les méthodes paramétriques permettent d'ajuster des distributions. La pertinence de ces méthodes et leur qualité de précision dépendent largement de la forme de distribution ainsi que l'espace de couleur choisi. Cependant, les phases d'apprentissage et de test pour ce type de méthodes sont lentes puisque celles-ci impliquent une procédure d'évaluation des paramètres utilisés. Parmi les avantages de ces méthodes, on peut citer le gain en espace mémoire ainsi qu'en possibilité de manipulation, plus d'intelligence dans la régularité des distributions et capacité d'interpoler les données d'apprentissage quand elles sont dispersées. La phase d'apprentissage peut être plus longue avec les approches paramétriques, mais celles-ci possèdent l'avantage de fournir une bonne estimation de la densité avec un ensemble d'apprentissage plus réduit [119]. Les meilleurs résultats offerts par ce type de méthodes est celui de Zhanyu Ma. [145]. Il représente un des meilleurs travaux en termes d'efficacité de détection avec celui de M. Abdullah-Al-Wadud et al. [4]. Cependant l'inconvénient que connait ce type de méthodes (paramétrique) est la forte complexité de calcul algorithmique durant l'estimation des paramètres [145].

Bien que les méthodes présentées précédemment utilisent la base de données Compaq de manières différentes pour définir les images/ pixels d'apprentissage et

les images/ pixels de test, et emploient différentes stratégies d'apprentissage, le tableau 4.6 donne une image assez fidèle des performances obtenues par ces méthodes. Une étude comparative est assez difficile à établir, mais néanmoins dans plusieurs travaux dont ceux cités sur le tableau précédent, des auteurs ont effectué une comparaison de leurs méthodes avec ceux de l'état de l'art malgré la diversité des stratégies d'apprentissage et de test utilisées [145] [76]. En analysant le tableau, on constate qu'il n'y a pas de type de méthode meilleur que les autres.

Notre travail est basé sur une méthode non paramétrique pour la détection de tout type de peau appartenant à différents races ethniques. Comme toutes les autres méthodes citées dans le tableau 4.6, notre méthode est dédiée à la détection de tout type de peau appartenant à différentes races ethniques. Etant donné que l'efficacité d'une approche paramétrique est tributaire de l'espace de couleur choisi, nous avons choisi d'exploiter plusieurs espaces de couleur. Nous présenterons par la suite les résultats de notre évaluation dans l'espace de couleur qui a donné les meilleurs résultats en comparaison avec ceux des méthodes présentées dans le tableau 4.6.

4.4 Notre approche

4.4.1 Hétérogénéité de la couleur de la peau

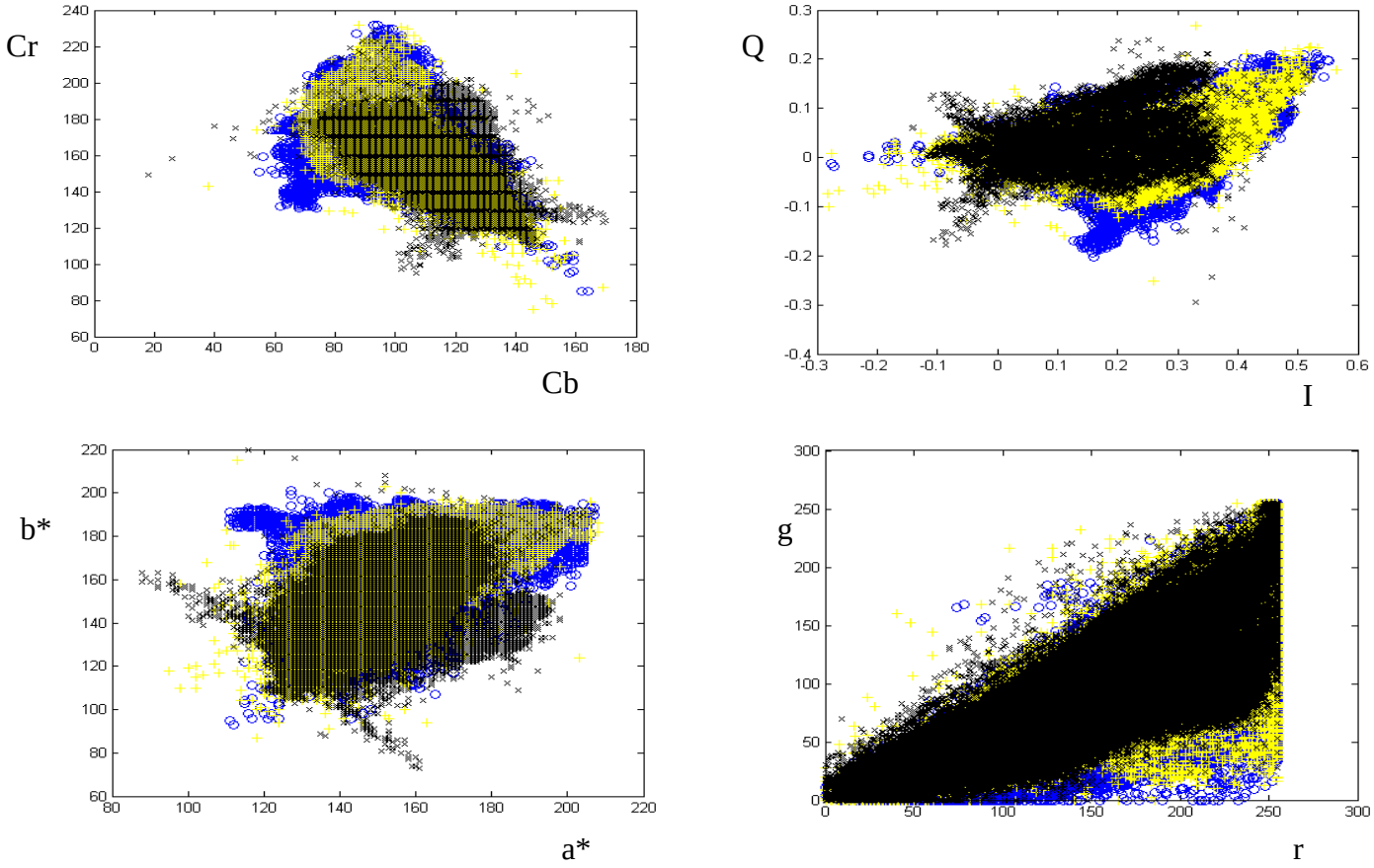
Toute la complexité de la détection de la peau émane de plusieurs facteurs à considérer. Parmi ces derniers, nous pouvons dénombrer les conditions d'acquisition de l'image, les conditions d'éclairage généralement inconnues, la variabilité de la teinte de la peau comme celle de tout objet en fonction du matériel utilisé pour l'acquisition. En outre, il existe un facteur important qui est la diversité ethnique avec des teintes qui varient selon la personne, la race et l'âge.

La couleur de la peau humaine appelée aussi teint ou complexion, présente une gradation continue du blanc au marron foncé presque noir, avec parfois des tons rosés ou cuivrés (voir figure 4.2). Classer les peaux depuis les plus clairs aux plus mates, revient à élaborer une taxonomie discriminative de l'espèce humaine. Selon le grand dictionnaire terminologique de l'office québécois de la langue française (http://www.oqlf.gouv.qc.ca), une race est un "Regroupement d'êtres humains, qui se distingueraient par des traits physiques communs héréditaires, généralement la couleur de leur peau, sans aucun égard à leur langue, à leur culture ou à leur pays d'origine". Selon le Compte Joseph Arthur de Gobineau [86], une race humaine est une subdivision de l'espèce humaine. Il divisait toutefois l'humanité en trois grandes races distinctes, la race blanche (aryenne), la race jaune et la race noire. Nous tenons compte donc de cette subdivision pour organiser la base d'images Compaq.

La figure 4.2 montre des exemples d'images de personnes d'origine ethnique distinctes avec des couleurs de peau différentes issues de la base d'images Compaq.

Figure. 4.2: La couleur de peau varie d'une personne à une autre.

La figure 4.3 montre la distribution colorimétrique des pixels de peau (extraits de la base d'image Compaq) appartenant à trois différentes races: blanche, jaune et noire dans les sous-espaces RG, IQ et a*b*.



+ : Race jaune o : Race blanche x : Race noire

Figure. 4.3: Densité de couleur de peau de différentes races dans les espaces CbCr, IQ, a*b* et rg.

Nous pouvons constater un recouvrement important des trois races considérés dans les quatre sous-espaces de couleurs présentés sur la figure 4.3. Cela peut nous laisser croire qu'un seul classifieur suffirait à la détection des pixels peau des trois races. En étudiant bien les graphes de près, nous pouvons constater des pixels isolées n'appartenant pas au recouvrement des trois races. Nous partons du principe que les performances de la régression logistique bayésienne à noyau se dégradent avec l'accentuation de l'hétérogénéité des pixels de la classe peau. En outre, ce type de régression est sensible aux pixels hors normes [200]. On sous-entend par *hors norme*, la particularité qu'ont des pixels à s'écarter notablement du nuage formé par des pixels appartenant à la classe dont la régression a pour objectif de discriminer. Néanmoins, ces éléments *isolés* sont difficilement classifiables à mesure que l'ensemble des pixels de la classe à laquelle ils appartiennent s'accroit lors de la phase d'apprentissage. D'après Viviane Planchon [166] "les valeurs des éléments hors normes dites *extrêmes* ne sont pas forcément des valeurs aberrantes. Une *valeur extrême* peut correspondre à un profil rare, intéressant à détecter et dont la suppression appauvrirait l'échantillon d'apprentissage et le modèle obtenu". Afin

d'améliorer les performances de détection, il serait intéressant de pouvoir mieux détecter les groupes de pixels isolés. Nous proposons une stratégie de classification visant à pouvoir détecter le plus possible de ces pixels. Au lieu d'élaborer un apprentissage d'un classifieur sur un ensemble contenant des pixels peau de toutes les races, nous opérons plusieurs apprentissages, chacun pour la détection de pixels d'une race bien définie. Cela nous mène donc à partitionner l'ensemble initial d'apprentissage en sous-ensembles de pixels dont la couleur correspond à l'une des races considérées et entrainer un classifieur à part par chacun d'eux. En d'autres termes, il est nécessaire d'avoir recours dans ce cas à plusieurs classifieurs, chacun consacré à la détection des pixels peau d'une race spécifique. Une stratégie de fusion des classifieurs qu'on abordera ultérieurement devrait alors être choisie pour la généralisation à la détection de tout type de peau.

4.4.2 Stratégies de classification multiple

Dans le cadre de notre travail, l'approche RLBN [131] est utilisée pour la détection d'un des sous-ensembles définis dans la section précédente. Détecter tous les pixels peau de différentes races revient à avoir recours à une régression RLBN multinomiale. La régression logistique multinomiale est une extension de la régression logistique pour le cas où il peut y avoir plusieurs variables explicatives. Celle-ci peut être implantée en utilisant plusieurs stratégies. Nous allons présenter les stratégies les plus célèbres et les plus utilisés qui sont L'approche "*un contre un*", l'approche "*un contre tous*" et l'approche "*ECOC*". Commençons par l'approche "*un contre un*".

Soit N_c le nombre de classifieurs considérés. Pour cette approche [100], il s'agit d'affecter à chaque couple de classes une fonction de décision. Cela revient à discriminer une classe d'une autre et donc, $N_c.\frac{N_c-1}{2}$ de ces fonctions seront apprises et chacune d'entre elles effectue un vote pour l'affectation d'un nouvel individu x à traiter. La classe de cet individu x devient ensuite la classe majoritaire après le vote. Cette approche est dite associée en général à une opération de vote majoritaire pour la fusion des décisions des sous-classifieurs pour la classification des nouveaux éléments [79].

L'approche "*un contre tous*" est la plus simple et la plus ancienne des méthodes de décomposition. Elle considère le problème comme un problème binaire multiple. Elle associe à chacune des classes considérées un classifieur qui lui est propre indépendamment des autres classes. Une fonction de décision est attribuée à chaque classifieur. Cette approche permet donc l'apprentissage de N_c de ces fonctions notées $\{f_i\}$, $\forall$ i = 1, …, N_c qui permettent de discriminer une classe parmi les autres avec N_c le nombre de classes considérées. Pour classer un individu, on le présente donc à tous les classifieurs et la décision s'obtient en appliquant le principe "*winners-takes-all*"; la classe retenue est celle associée au classifieur ayant renvoyé la valeur la plus élevée.

Si la fonction $f_i(x)$ correspondant au $i^{ième}$ classifieur est maximale alors le nouveau individu x sera affecté à la classe C_i [180], avec

$$Ci = argmax \{f_i(x) / i = 1, \ldots, N_c\} \quad (4.8)$$

L'approche *ECOC* (Error Correcting Output Coding) quand à elle combine les résultats obtenus à partir des classifieurs binaires. La classe d'un individu x est celle qui pour les différents classifieurs binaires donne le moins d'erreurs [63]. Elle est basée sur la matrice de codes ($M_{i,j}$, i = 1, …, N_c, j = 1, …, n_c). Celle-ci représente la contribution de chaque classifieur au résultat final de la classification en se fondant sur les erreurs commises par les différents classifieurs.

La matrice de codes est une matrice de taille N_c x n_c où N_c est le nombre de classes et en colonne n_c classifieurs. Elle représente la contribution de chaque classifieur au résultat final de la classification en se basant sur le résultat des erreurs commises par les différents classifieurs. La matrice de codes sert à tester les données dans différents cas (base de test connue et inconnue):

- Dans le cas où la base de test est connue, chaque classifieur S_j donne la distance de chaque point de la base de test par rapport à l'hyperplan du classifieur. Ainsi, la classe d'un individu x est donnée par:

$$classe(x) = \arg\min_i \left(\sum_{j=1}^{n_c} (1 - d_j(x)\, M_{i,j}) \right) \tag{4.9}$$

où $d_j(x)$ est la distance entre x et l'hyperplan du classifieur S_j .

- Dans le cas où la base de test est inconnue, chaque classifieur S_j prédit la classe d'un individu soit positive (+1) ou négative (-1). Le résultat est donc un vecteur de taille n_c. Ce vecteur est comparé par la suite à chaque ligne de la matrice des codes en utilisant la distance de Hamming. La ligne garantissant la plus petite distance est la classe assignée à l'individu x. Cette technique est dite de *décodage avec la distance de Hamming* [109].

4.4.3 Stratégie adoptée pour l'application de notre approche multinomiale

L'approche "*Un contre tous*" correspond à l'approche de classification multiple la plus utilisée. Celle-ci développe K ensembles, en opposant chaque classe k aux données à traiter. On obtient donc K modèles de prédiction qui, pour un pixel donné, vont chacun estimer la probabilité qu'un pixel appartienne à la classe k [198]. En d'autres termes, cela revient à élaborer plusieurs classifieurs, chacun est attribué à une classe de manière à la discriminer de toutes les autres. Dans notre cas, nous traitons un problème de classification multiple puisqu' un pixel peut appartenir soit à l'une des races considérées soit à la classe non peau. Nous nous sommes inspirés de la stratégie "*Un contre tous*" avec une légère modification sur le nombre de modèles de prédiction à utiliser. Pour K classes, Nous nous sommes contentés de K-1 classifieurs. Nous attribueront un modèle de prédiction pour chaque classe relative à une des races considérées et nous omettons d'établir un autre pour la classe non peau. Si aucun de ces modèles ne reconnait un pixel comme appartenant à la classe lui correspondant alors celui-ci est automatiquement affecté à la classe non peau. En d'autres termes, nous avons jugé inutile d'affecter un classifieur supplémentaire à la classe non peau puisqu'il ne peut y avoir des pixels qui n'appartiennent pas simultanément aux classes peau et non peau. Il suffit qu'un pixel soit rejeté par tous les classifieurs de peau pour qu'il soit considéré comme non peau.

A cet effet, nous avons pris en considération une mixture de trois régressions logistiques bayésiennes à noyau, chacune correspondant à une des plusieurs races traitées. Nous avons donc construit trois classifieurs, chacun est consacré à la détection de pixels peau d'une des trois races: race noire, race blanche et race jaune puis nous avons défini une stratégie de fusion des sorties de ces classifieurs qu'on présentera ultérieurement pour la détection de tout type de peau.

La section 4.4.4 présentera l'approche RLBN [131] pour la détection de la peau d'une manière générale. Nous aborderont dans la section 4.5 la stratégie qu'on a adoptée afin d'améliorer les performances de détection de notre classifieur dans le cas de détection de tout type de peau.

4.4.4 Modèle de prédiction

Dans le cadre de l'apprentissage supervisé, nous supposons que l'échantillon d'apprentissage est représenté sous la forme: $\left\{ \tilde{X}, Y \right\}$ où: $\tilde{X} = \{X_i\}_{i=1}^{N}$ avec $X_i \in R^k$ *(k≥1)* $\forall\ i \in \{1,2\ldots..,N\}$ sont les variables d'entrée correspondant aux coordonnées des pixels dans l'espace de couleur de dimension k choisi pour la présentation et $Y \in [0,1]$ la variable binaire de sortie correspondant à la classe considérée. Il s'agit de la probabilité d'appartenance à la classe peau dans notre cas. L'objectif est de déterminer à partir de cet ensemble la liaison qui associe à chaque variable d'entrée une variable de sortie. Une fois entrainé, notre classifieur possède alors la faculté d'établir des prédictions pertinentes sur les variables de sortie correspondantes à de nouvelles variables d'entrée. La régression logistique classique permet d'établir un classifieur linéaire dans l'espace de couleur de présentation permettant de discriminer les pixels peau de ceux de non peau. Afin d'améliorer les performances de cette classification, la méthode à *noyau* permet d'effectuer une projection de la représentation des pixels de l'espace de couleur (espace de représentation d'origine) vers un espace de dimension généralement plus grand *F* connu sous le nom *d'espace de caractéristiques*. Cette projection permet de mettre en évidence les différences entre les valeurs portées par les pixels d'une manière significative. La distribution des pixels dans cet espace est plus éparse que dans l'espace d'origine et la distinction entre les pixels peau et non peau est facilitée. Cela permet d'améliorer remarquablement le processus de séparation entre les pixels peau et les pixels non peau. Nous notons que dans l'espace de caractéristiques, la discrimination d'une classe par rapport à une autre se fait d'une manière linéaire. La fonction qui assure cette conversion est présentée comme suit [131]:

$$\Phi(X) = (1, \phi_1(X), \phi_2(X), ..., \phi_N(X)) \in F \qquad (4.10)$$

où: X est un point correspondant à la représentation d'un pixel dans l'espace de couleur utilisé. $(1, \phi_1(X), \phi_2(X), ..., \phi_N(X))$ correspond à la représentation de X dans l'espace *de caractéristiques* (*Features space*) de dimension N.
$\phi_i(X) = K(X, X_i)$ représente la fonction noyau correspondant à la coordonnée de X dans l'axe i+1 du nouvel espace de caractéristiques [59].
Ce modèle est plus connu sous le nom de *l'astuce du noyau*. Il a été publié en 1964 par M. Aizerman et al. [9]. (Pour plus de détails, voir la section 3.9.1 du chapitre 3).

Le modèle de prédiction qui permet de discriminer la classe à considérée dans l'espace des caractéristiques est présenté par la fonction *f(X,w)* comme suit:

$$f(X;w) = \sum_{i=1}^{N} w_i \phi_i(X) + w_0 = w^T \Phi(X) \tag{4.11}$$

où: N est la taille de l'échantillon d'apprentissage et $f(X;w)$ représente une fonction logarithmique qu'on présente comme suit:

$$f(X;w) = \log(\frac{Y}{1-Y}) \tag{4.12}$$

En remplaçant (4.12) dans (4.11), nous obtenons:

$$\log(\frac{Y}{1-Y}) = \sum_{i=1}^{N} w_i \phi_i(X) + w_0 = w^T \Phi(X) \quad \Rightarrow$$

$$Y = \frac{\exp(w^T \Phi(X))}{1 + \exp(w^T \Phi(X))} \tag{4.13}$$

Afin d'établir un tel modèle de prédiction, il est nécessaire de pouvoir estimer les paramètres w_i afin d'effectuer une discrimination des deux classes peau et non peu d'une manière optimale. Tous les détails sur l'estimation de ces paramètres et l'algorithme correspondant sont présentés dans l'annexe.
La classification finale d'un nouveau pixel X à traiter se fait de la manière suivante:

$$\begin{cases} Si\ Y > \lambda \quad \text{alors} \quad X \in peau \\ Sinon \qquad\qquad X \in non\ peau \end{cases} \tag{4.14}$$

avec λ un seuil séparateur entre les deux classes peau et non peau à fixer expérimentalement le plus souvent ajusté à 0.5.
Il a été montré dans [131] que la régression logistique bayésienne à noyaux donne de meilleurs résultats de classification que toutes les autres méthodes basées sur les noyaux de Fisher comme les séparateurs à vaste marge (*SVM*), la méthode adaboost régularisé (AB_R), les machines à vecteurs de pertinence (*RVM*), etc.
Le modèle de prédiction que nous venons de présenter sert à la construction d'un seul classifieur. Celui-ci peut servir à la détection des pixels peau d'une seule race si la phase d'apprentissage est effectuée exclusivement sur des pixels correspondant à cette race. Il peut servir également à la détection de tout type de peau à condition que la phase d'apprentissage soit effectuée sur toutes les races à la fois de manière équilibrée.

En régression logistique, l'apprentissage des données doivent se faire de manière rationnelle pour que le classifieur puisse agir de la meilleure manière. Les différentes couleurs de peau d'une même race doivent donc être présentes dans la base

d'apprentissage d'une manière équitable. Toute représentation faible d'une certaine couleur dans le corpus entraine sa négligence par le classifieur au profit des autres couleurs. Il y'a lieu donc de faire un équilibre entre les différentes représentations chromatique. Autrement dit, les ensembles de pixels peau de chaque race doivent être approximativement de même taille. Nous présenterons les performances de détection d'un tel classifieur ultérieurement.

Afin d'améliorer les performances de détection dans le cas de détection de tout type de peau, plusieurs classifieurs basés sur le modèle que nous venons de présenter peuvent être exploités comme montré dans la section 4.4.1. La stratégie d'utilisation de ces derniers est présentée dans la section suivante.

4.4.5 Combinaison de plusieurs classifieurs de peau

4.4.5.1 Introduction

La régression présentée par l'équation 4.11 permet bien de séparer les pixels peau de ceux de non peau dans l'espace de caractéristiques d'une manière linéaire. Pour des raisons que nous avons évoqué préalablement, il serait préférable de traiter chaque type de peau correspondant à une race ethnique à part. Comme expliqué dans la section 4.5.1, nous avons utilisé une régression RLBN multinomiale pour la détection de tout type de peau. A cet effet, plutôt que de considérer un cluster correspondant à la couleur peau, nous nous intéressons à définir trois types de classes, chacune étant spécifique à l'une des trois races humaines: la race blanche, la race noire et la race jaune. Nous traitons donc dans ce cas un problème de classification multiple. A cet effet, il est nécessaire de prendre en considération trois types de régression de la forme de l'équation 4.11, chacune traitant un problème de classification de type binaire. Une fois la classification effectuée, une fusion des trois classes de peau s'avère nécessaire pour distinguer la peau de la non peau. Il s'agit alors de définir une mixture de trois régressions logistiques bayésiennes à noyau réalisée par fusion de trois classifieurs de base.

4.4.5.2 Stratégies de fusion de plusieurs classifieurs

Il existe plusieurs règles de fusion des résultats des classifieurs. Les auteurs dans [66], ont utilisé des règles de fusion de plusieurs classifieurs dans le cadre de la détection de la peau. Parmi les stratégies qu'ils ont utilisée il y'a la règle du "*vote majoritaire"*, le "*ou logique*", le "*et logique*" et la "*pondération des entrée et l'addition des sorties*". Cette dernière s'appuie sur l'affectation d'un poids fixé expérimentalement à chaque classifieur puis une addition du résultat de chacun est opérée pour une décision finale. Les auteurs ont montré la supériorité de cette stratégie sur toutes les autres. Nous pouvons noter aussi que les auteurs dans [143] ont défini six stratégies de fusion de classifieurs: le "*minimum"*, le "*maximum"*, la "*moyenne"*, la "*médiane"*, le "*vote majoritaire"* et la stratégie *"Oracle"*. Ils ont montré la supériorité de la stratégie Oracle sur toutes les autres. Le principe de cette dernière est que si au moins un des classifieurs considère que la donnée traitée est de la classe X alors tous les autres classifieurs la considèrent comme telle. Elle correspond à la stratégie du "*ou logique*" définie dans [66].

Nous avons opté pour la stratégie "Oracle" ou du "ou logique" comme l'une des règles de fusion de nos classifieurs. Cela se justifie par le fait que cette stratégie semble être la plus adéquate dans le cas de notre étude. Vu la diversité des couleurs de peau, un pixel peau d'une race donnée pourrait très bien être détecté par un classifieur relatif à cette race sans pour autant être détecté par les autres. En d'autres termes, il suffit qu'un seul classifieur relatif à un type de peau détecte le pixel à traiter comme peau pour le considérer comme tel. Comme la structure de notre système est fondée sur la base de la fusion de plusieurs classifieurs de type RLBN simple, le taux de fausse détection apporté par chacun de ces derniers contribue à la dégradation des performances de détection. Afin de remédier à ce problème, nous avons définis une autre stratégie de type "minimum". Vu le recouvrement important entre les distributions colorimétriques des pixels peau des différentes races présenté sur la figure 4.3 concernant les sous-espace colorimétriques considérés, même si un pixel est détecté comme peau par un seul classifieur, la probabilité d'être détecté comme tel par les autres ne doit pas être au dessous d'un certain seuil fixé expérimentalement que nous noterons *S*. Tenant compte du chevauchement entre les classes de pixels des différentes races, aucun classifieur ne produirait une probabilité *très faible* pour un pixel peau quelque soit la race à laquelle il correspond même s'il ne le détecte pas comme peau. Il suffirait alors que la probabilité produite par un seul classifieur soit au dessous du seuil *S* pour que le pixel traité soit considéré comme non peau malgré qu'il existe d'autres classifieurs qui considèrent ce pixel comme peau. A cet effet, nous avons considéré le *minimum* des probabilités produites par tous les classifieurs de base. Il suffit alors que la valeur de ce minimum soit inférieure à *S* pour considérer le pixel traité comme non peau. Cette condition est représentée comme suit:

$$Si \ \min(P_1, P_2, ..., P_i) < S \ \text{ alors } X \in non \ \ peau \tag{4.15}$$

où: P_i est la probabilité produite par le classifieur i selon l'équation 4.13 et X un pixel à traiter.

Cette stratégie de type "minimum" permet de remédier efficacement au problème du taux élevé de fausse détection causée par la fusion des classifieurs de base.

Nous avons donc étendu notre méthode de telle manière à pouvoir détecter les pixels peau quelque soit la race ethnique à laquelle ils correspondent d'une manière plus efficace. Nous avons donc consacré pour chaque race un ensemble d'images qui va servir de base d'apprentissage pour le classifieur qui lui est approprié. Une fois entrainés, tous les classifieurs sont impliqués dans la classification des pixels de peau et de non peau d'une image à tester. Comme montré dans la section précédente, chaque classifieur produit une probabilité d'appartenance à la classe peau. Nous avons donc définis un seuil pour chaque classifieur permettant la séparation entre la classe peau et la classe non peau qu'on notera S_i. Ces seuils sont fixés expérimentalement durant la phase de test de manière à effectuer une bonne classification des pixels.

La figure 4.4 illustre le principe de fonctionnement avec Y_i un vecteur correspondant à une fonction de sortie indiquant l'appartenance de chaque pixel de l'image traitée à l'une des deux classes considérées. Si Y_i [j]= 1, la classe d'affectation est la classe peau, sinon si Y_i [j]= 0 alors la classe d'affectation est la classe non peau avec j un indice référant le $j^{ième}$ pixel de l'image. Nous notons que Y est un vecteur de même type que Y_i. P_j représente un vecteur correspondant aux probabilités d'appartenance de chaque pixel à la classe peau calculé par le classifieur j et S_j correspond au seuil de séparation des deux classes peau et non peau utilisé par le même classifieur. S représente un seuil qui permet de réduire le taux de fausse détection comme expliqué précédemment. Les vecteurs cités comme arguments dans les fonctions *MIN* et *OR* sont considérés comme une matrice conçue à partir d'une fusion horizontale effectuée sur ces vecteurs. Les fonctions *Min* et *OR* opèrent ligne par ligne sur cette matrice sachant que chaque ligne correspond à un pixel et chaque colonne à un classifieur. Le produit d'une telle fonction est donc un vecteur de même taille que ceux définis comme arguments (la taille de l'image traitée).

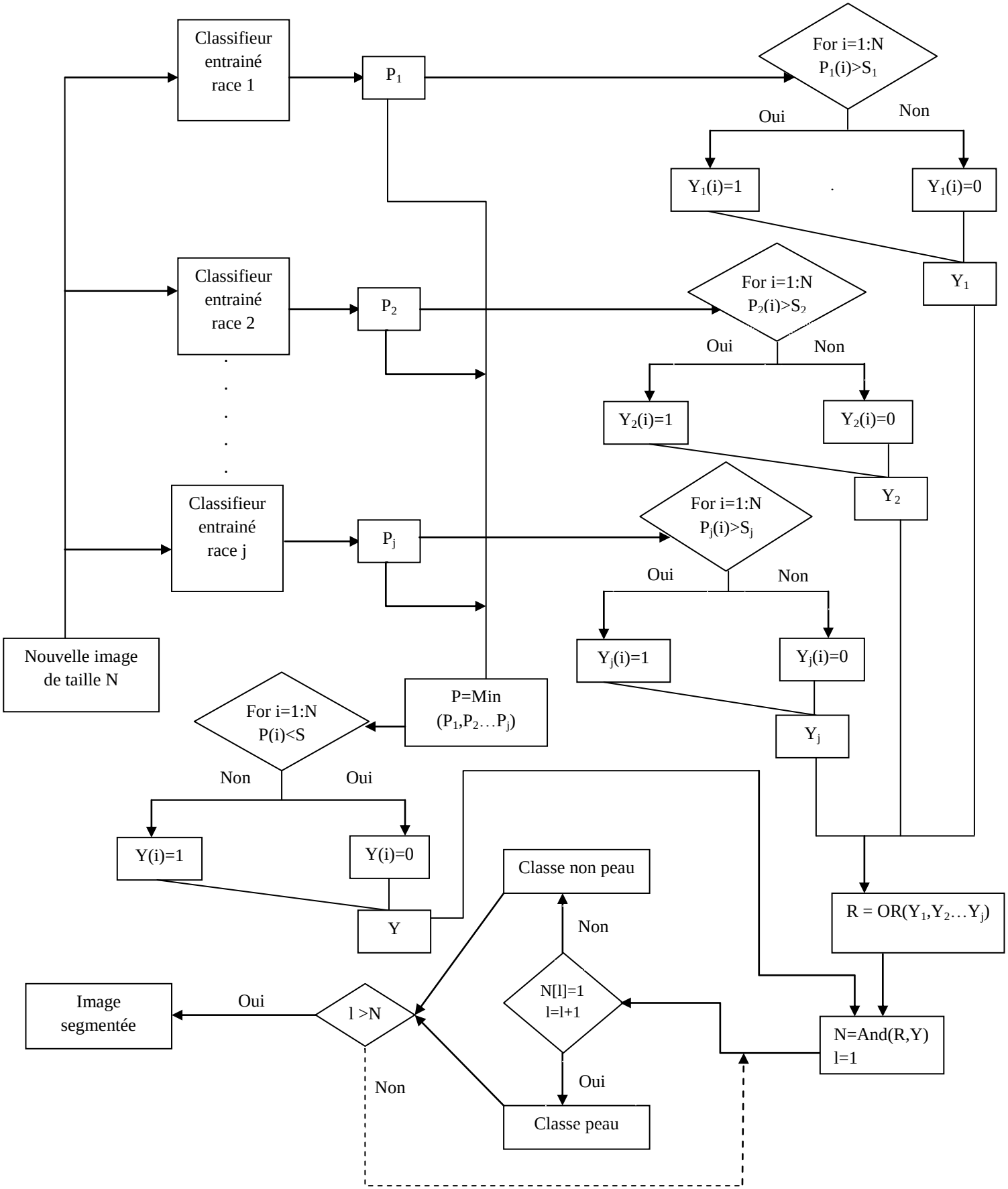


Figure. 4.4: Combinaison de plusieurs classifieurs pour la détection de tout type de pixels peau.

4.5 Résultats expérimentaux

4.5.1 Environnement de développement (Matlab)

Le nom de *Matlab* est dérivé du terme Matrix Laboratory. Il s'agit d'un outil de développement très puissant riche en beaucoup d'outils intégrés destinés à la résolution de problèmes scientifiques et, plus généralement, pour tous les domaines où des calculs numériques importants doivent être faits. C'est un système interactif de calcul numérique utilisable comme une calculette et qui dispose d'un grand nombre de fonctions, d'un langage de programmation et d'outils de visualisation graphique. La structure de base sur laquelle travaille Matlab est la matrice [11]. Cela se fait sans déclarations de type et les allocations de mémoire ou les redimensionnements sont effectués sans intervention de l'utilisateur, d'où une grande souplesse d'emploi. Matlab met à disposition de l'utilisateur plusieurs fenêtres (voir figure 4.5). Parmi ces fenêtres, nous pouvons distinguer:

- La fenêtre de commande (*command window*) où l'on exécute des fonctions Matlab directement où des fichiers créé par l'utilisateur.

- La fenêtre de lancement (*launch pad*), où l'on peut voir l'arborescence du dossier Matlab et appeler aisément la documentation, les démonstrations et les fichiers scripts.

- La fenêtre de l'espace de travail (*workspace*) où sont visualisées toutes les variables utilisées dans la session ouverte (taille pour les matrices, valeurs pour les constantes).

- La fenêtre du répertoire courant (*current directory)* qui permet de visualiser des fichiers et matrices stockées dans le répertoire.

D'autres fenêtres peuvent être générées à partir de l'environnement Matlab comme la fenêtre de l'aide et la fenêtre des fichiers scripts. Toutes ces fenêtres peuvent être ouvertes ou fermées selon le besoin de l'utilisateur.

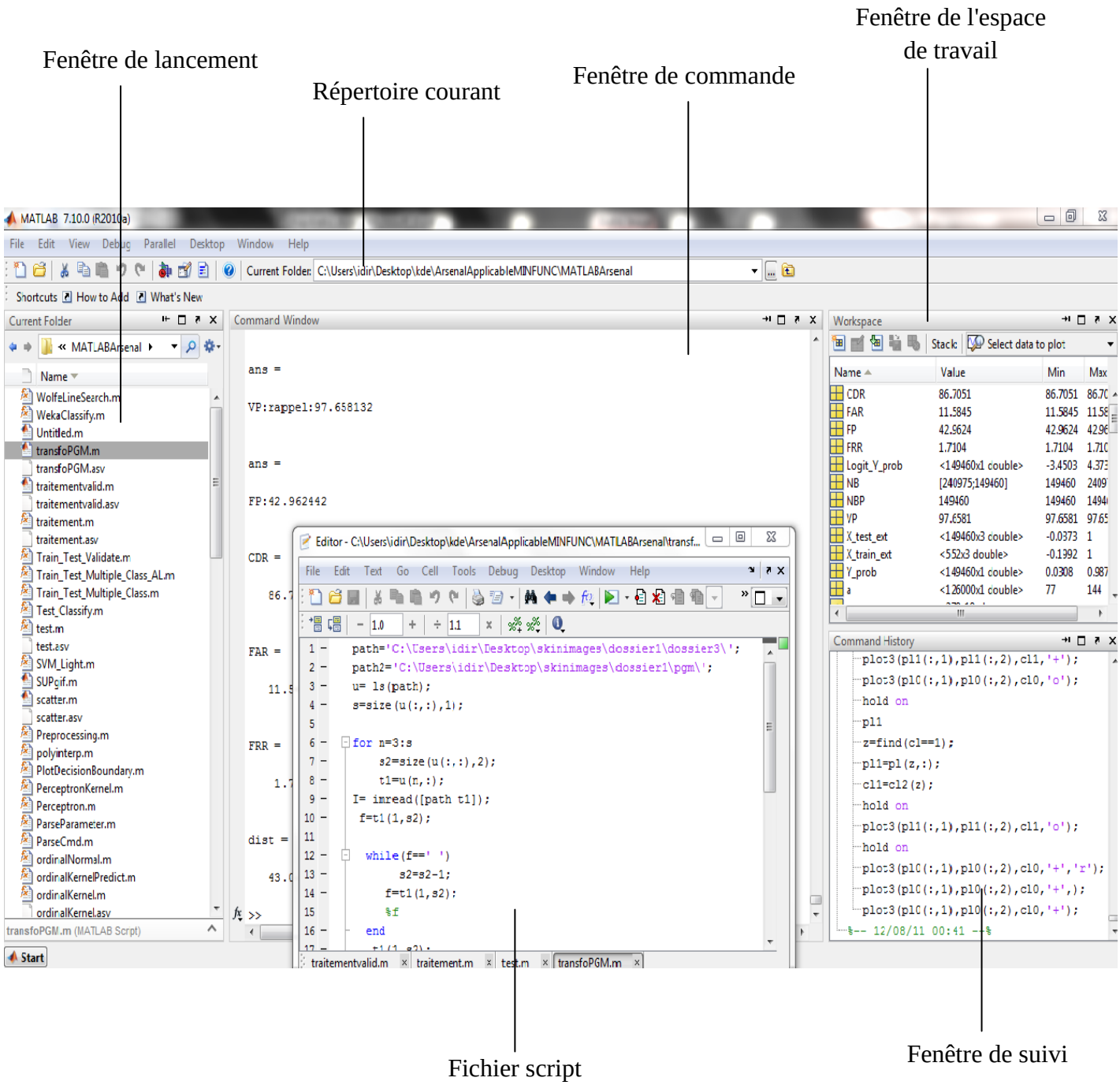


Figure. 4.5 Environnement de développement Matlab.

Etant un langage initialement créé pour traiter des problèmes d'algèbres linéaires, donc, en tant qu'outil optimisé pour le calcul matriciel, Matlab convient parfaitement pour le développement rapide d'algorithmes et de programmes pour la manipulation d'images numériques. Ces dernières sont en effet représentées par des matrices à 2 ou 3 dimensions.

Matlab fonctionne suivant deux modes, le mode interactif et le mode programmation. Dans les deux cas, l'utilisateur peut définir ses propres fonctions et les utiliser.

Le mode interactif correspond au mode calculette. Les instructions sont tapées successivement dans l'ordre d'exécution dans la fenêtre de commande [11]. Chaque ligne d'instructions est terminée par un retour chariot qui valide la ligne. La ligne d'instructions est alors interprétée et les instructions qu'elle contient sont immédiatement exécutées. Le résultat (respectivement un message d'erreur) est affiché en cas de bon fonctionnement (respectivement en cas d'erreur). L'affichage du résultat d'une instruction est parfois indésirable: pour le supprimer, il suffit de terminer l'instruction par un point-virgule.

A la différence du mode interactif, les instructions dans le mode programmation sont tapées successivement dans l'ordre d'exécution dans un fichier dont l'extension (fin du nom du fichier) est *.m* [11], par exemple essai.m, et cela se fait en utilisant un éditeur de texte. Pour faire exécuter les instructions, il suffit alors de taper le nom du fichier sans l'extension (essai dans le cas de l'exemple) dans la fenêtre de commandes. L'avantage de cette façon de travailler est qu'en cas d'erreur dans l'une des instructions, il n'est pas nécessaire de tout retaper: il suffit de corriger l'erreur dans le fichier, de sauvegarder le fichier et de relancer la commande. On désigne un tel fichier du nom de script. Quand la commande essai est lancée, Matlab cherche le fichier essai.m dans le répertoire de travail. Il faut lui avoir préalablement indiqué le chemin qui mène à ce répertoire.

4.5.2 Evaluation expérimentale

4.5.2.1 Stratégie d'évaluation

Nous avons effectué nos tests expérimentaux sur la base d'images Compaq décrite dans la section 4.2. Nous avons d'abord opéré une évaluation basée sur l'apprentissage d'un seul classifieur. Nous avons par la suite opéré une évaluation combinant les trois classifieurs chacun dédié à la détection d'un type de peau spécifique formant ainsi un classifieur de synthèse pour la détection de la couleur de tout type de peau.

Les méthodes présentées sur le tableau 4.6 utilisent différentes stratégies d'apprentissage et de test. Une étude comparative est assez difficile à établir, mais néanmoins dans plusieurs travaux dont ceux cités sur ce tableau, des auteurs ont effectué une comparaison de leurs méthodes avec ceux de l'état de l'art malgré la diversité des stratégies d'apprentissage et de test utilisées [145] [76]. Durant l'évaluation, nous avons entrainé chaque classifieur avec une collection d'environ une centaine d'images choisies d'une manière aléatoire, chacune contenant des pixels de peau correspondant à la race pour laquelle le classifieur est destiné et des pixels non peau. Comme indiqué dans la section 4.4.4, ces collections doivent être approximativement de même taille.

Nous avons construit trois classifieurs de base. Chacun est destiné à la détection des pixels peau d'une des trois races: blanche, jaune et noire respectivement. Nous avons attribué au premier classifieur une collection d'images contenant des pixels peau de race blanche, au second une collection d'images contenant des pixels peau de race jaune et au troisième une autre collection contenant des pixels peau de race noire. Chacune de ces trois collections est constituée d'une centaine d'images qui

serviront d'apprentissage pour chaque classifieurs. Pour des fins de test, nous avons attribué à chaque classifieur une collection d'images contenant des pixels peau correspondant à la race auquel le classifieur est destiné. La taille de chaque collection de test est d'environ une centaine d'images. Le choix des images de test et d'apprentissage se fait d'une manière aléatoire.

A partir de la fusion des trois classifieurs de base, nous avons construit un autre classifieur destiné à la détection des pixels peau de toutes les races comme indiqué dans la section 4.4.5.2. Pour des besoins de comparaison, nous avons également utilisé un autre classifieur de base entrainé à la fois sur des pixels peau de toutes les races considérées à savoir la race blanche, jaune et noire. Nous attribuons à celui-ci une collection d'apprentissage construite à partir de toutes les images qui constituent les trois ensembles d'entrainement décrits préalablement. Nous définissons en plus une autre collection d'images de test constituée de toutes les images appartenant aux trois ensembles de test définis plus haut. En termes de nombre de pixels, cela nous fait 34290375 pixels pour l'apprentissage et 46086886 pour le test.

4.5.2.2 Nouvelle mesure d'évaluation

Il est difficile d'effectuer une comparaison entre les différents classifieurs en utilisant directement les critères d'évaluation VP (*Vrais positifs*) et FP (*Faux positifs*) que nous avons décris dans la section 4.3.2 surtout quand les scores à comparer sont assez proches les uns des autres. Dans [145], les auteurs ont considéré qu'un classifieur parfait agirait avec une valeur de VP qui s'élève à 100% et une valeur nulle de FP. Cette valeur est présentée sous la forme d'un point de coordonnées (0,100) dans les graphes des figures 4.6 et 4.7. Les auteurs affirment que plus le point correspondant au résultat d'un classifieur donné est proche du point de coordonnées (0,100) plus ses performance sont meilleurs. A cet effet, nous avons défini un nouveau critère d'évaluation qui se base sur les mesures de VP et FP et offre une facilité de comparaison remarquable entre les classifieurs. Il s'agit de la distance euclidienne d'un point relatif à un résultat d'une expérimentation au point de coordonnées (0,100). Ce point est appelé *point optimal*. Plus cette distance est petite, plus le classifieur est meilleur [145]. Elle est calculée selon l'équation suivante:

$$d = \sqrt{(100 - vp)^2 + FP^2} \tag{4.16}$$

Nous avons également utilisé les critères d'évaluations CDR, FAR et FRR décris dans la section 4.3.2 afin de comparer notre méthode avec celle présentée dans [66]. Nous présentons dans les tableaux de la section suivante les résultats de nos évaluations pour chaque classifieur dans divers espaces de couleurs à savoir IQ, YIQ, CbCr, YCbCr, xyz et rgb.

4.5.2.3 Résultats expérimentaux

Nous avons effectué une évaluation sur chaque race à part puis sur toutes les races confondues conformément à la stratégie présentée dans la section 4.5.2.1. Nous présentons les résultats de nos évaluations dans les tableaux suivants. Nous notons que les valeurs présentées dans tous ces tableaux correspondent à des pourcentages. Les valeurs représentées en gras correspondent aux meilleurs résultats de chaque mesure d'évaluation dans chaque espace de couleur de conversion utilisé.

Tableau. 4.7: Résultats expérimentaux obtenus sur des images contenant des pixels peau de race blanche.

	Espace de couleur	VP	FP	Distance-point optimal	CDR	FAR	FRR
RLBN multinomiale	IQ	92.894214	9.228108	**11.6469**	**91.3794**	6.5866	2.0340
	YIQ	79.904707	**4.706585**	20.6391	90.8884	**3.3593**	5.7523
	CbCr	93.375906	9.67411	11.7246	91.199	6.9049	1.8962
	YCbCr	78.076885	4.797360	90.3004	3.4241	6.2755	22.4419
	xyz	**94.384434**	10.379414	11.8011	90.9842	7.4083	**1.6075**
	rgb	88.241335	8.69098	14.6219	90.4414	6.2329	3.3257
RLBN simple	IQ	94.906239	10.897732	12.0294	90.7637	7.7782	1.4521
	YIQ	94.858808	**10.857251**	**12.013**	**90.7790**	**7.7494**	1.4717
	CbCr	95.003681	11.008355	12.0891	90.7126	7.8572	1.4302
	YCbCr	94.934397	10.987249	12.0988	90.7078	7.8421	1.45
	xyz	**98.144935**	26.400309	26.4654	80.6258	18.8432	**0.5310**
	rgb	96.02018	23.330072	23.6671	82.209	16.6518	1.1392
RLBN race blanche	IQ	91.837602	8.849065	12.0387	91.3475	6.3160	2.3365
	YIQ	89.197313	10.701287	15.2058	89.2697	7.6380	3.0923
	CbCr	91.859096	**8.807339**	**11.9935**	**91.3834**	**6.2862**	2.3303
	YCbCr	89.114244	10.874531	15.3868	89.1223	7.7617	3.1161
	xyz	**97.992548**	24.322531	24.4052	82.0652	17.3602	**0.5746**
	rgb	91.784814	9.675368	12.6926	90.7426	6.9058	2.3516

Tableau. 4.8: Résultats expérimentaux obtenus sur des images contenant des pixels peau de race jaune.

	Espace de couleur	VP	FP	Distance-point optimal	CDR	FAR	FRR
RLBN multinomiale	IQ	94.091166	10.883777	**12.3843**	90.2958	8.3032	1.401
	YIQ	79.643283	**4.969958**	20.9546	**91.3818**	**3.7916**	4.8267
	CbCr	**94.373377**	12.415069	13.6303	89.1245	9.4714	**1.3341**
	YCbCr	82.450661	7.023394	18.9026	90.4809	5.3581	4.161
	xyz	94.250028	13.287096	14.4779	88.5	10.1367	1.3633
	rgb	88.44304	9.360808	14.8723	90.1185	7.1413	2.7402
RLBN simple	IQ	95.852193	**13.073543**	**13.7158**	**89.0428**	**9.9738**	0.9835
	YIQ	95.911475	13.133896	13.7556	89.0108	10.0198	0.9694
	CbCr	95.846662	13.861153	14.47	88.4406	10.5746	0.9848
	YCbCr	95.846352	13.798627	14.4102	88.4882	10.5269	0.9848
	xyz	**97.172091**	31.923996	32.0491	74.9746	24.3547	**0.6707**
	rgb	94.220076	27.646445	28.2442	77.5382	21.0914	1.3704
RLBN race jaune	IQ	94.259194	**11.102215**	**12.4986**	**90.1690**	**8.4698**	1.3612
	YIQ	85.653934	29.253714	32.582	74.281	22.3175	3.4015
	CbCr	94.431789	12.446443	13.6352	89.1844	9.4953	1.3202

	YCbCr	87.06306	28.445941	31.2496	75.2313	21.7013	3.0674
	xyz	**98.39625**	33.379832	33.4183	74.1544	25.4653	**0.3803**
	rgb	96.891925	30.155801	30.3155	76.2573	23.0057	0.7369

Tableau. 4.9: Résultats expérimentaux effectués sur des images contenant des pixels peau de race noire.

	Espace de couleur	VP	FP	Distance-point optimal	CDR	FAR	FRR
RLBN multinomiale	IQ	94.274359	9.158964	**10.8014**	91.426	7.5984	0.9756
	YIQ	96.01056	27.331610	27.6212	76.6456	22.6747	0.6797
	CbCr	93.702334	9.317394	11.2461	91.1971	7.7298	1.073
	YCbCr	**96.881431**	28.376522	28.5474	75.9271	23.5416	**0.5314**
	xyz	91.368439	**8.26745**	11.9522	**91.6705**	**6.8588**	1.4707
	rgb	95.050719	17.875136	18.5477	84.3273	14.8295	0.8433
RLBN simple	IQ	92.7398	8.288808	**11.0188**	**91.8865**	6.8765	1.237
	YIQ	92.29065	**8.234979**	11.2805	91.8546	**6.8319**	1.3136
	CbCr	92.089231	8.577212	11.6683	91.5363	7.1158	1.3479
	YCbCr	91.553669	8.47276	11.9636	91.5317	7.0291	1.4391
	xyz	**97.478542**	22.991137	23.129	80.4966	19.0738	**0.4296**
	rgb	96.569458	18.473392	18.7892	84.0897	15.3258	0.5845
RLBN race noire	IQ	96.99534	**11.539198**	**11.924**	**89.915**	**9.5731**	0.512
	YIQ	92.496445	23.707595	24.8667	79.0534	19.6681	1.2785
	CbCr	96.802672	12.634763	13.033	88.9732	10.482	0.5448
	YCbCr	92.904388	22.683412	23.7673	79.9725	18.8185	1.209
	xyz	**97.025677**	17.996213	18.2403	84.5633	14.9299	**0.5068**
	rgb	93.070657	21.749393	22.8266	80.7757	18.0436	1.1807

Tableau. 4.10: Résultats expérimentaux obtenus sur des images contenant des pixels peau de toutes les races.

	Espace de couleur	VP	FP	Distance-point optimal	CDR	FAR	FRR
RLBN multinomiale	IQ	94.305598	11.776739	**13.0812**	**89.9872**	8.3614	1.6514
	YIQ	71.0435	**5.582151**	29.4896	87.6391	**3.9633**	8.3977
	CbCr	94.508405	12.888082	14.0093	89.257	9.1504	1.5926
	YCbCr	73.352730	5.915414	27.2960	88.0721	4.1999	7.728
	xyz	**95.094543**	14.31027	15.1277	88.4172	10.1601	**1.4226**
	rgb	88.473663	9.742908	15.0924	89.7399	6.9174	3.3428
RLBN simple	IQ	96.36617	14.272677	14.728	88.8127	10.1335	1.0538
	YIQ	96.240371	**14.090461**	**14.5834**	**88.9056**	**10.0041**	1.0903

	CbCr	96.118485	14.67735	15.1819	88.4536	10.4208	1.1257
	YCbCr	95.957072	14.523958	15.0762	88.5156	10.3119	1.1725
	xyz	**98.090001**	34.585872	34.638	74.8909	24.5552	**0.5539**
	rgb	96.371989	33.841112	34.035	74.921	24.0269	1.0522

Nous pouvons remarquer que dans chaque cas étudié, le classifieur basé sur la RLBN multinomiale agit mieux que tous les autres dans les espaces de couleur IQ, CbCr, xyz et rgb en assurant une distance minimale au point optimal. Néanmoins, nous avons constaté que les performances de détection de ce classifieur se dégradent considérablement dans les espaces de couleur pourvus de la composante de luminance comme confirmé dans [184]. Par contre, les classifieurs basés sur une RLBN simple sont beaucoup plus stables dans ces espaces. Il arrive même que la détection soit plus performante en tenant compte de la composante de luminance comme dans le cas présenté dans le tableau 4.10. Les meilleurs résultats obtenus sont ceux de la RLBN multinomiale effectués dans l'espace de couleur IQ suivis de ceux effectués dans l'espace CbCr.

Afin de montrer l'utilité de la stratégie "*Min*", nous effectuons une évaluation du classifieur basé sur une RLBN multinomiale sans avoir recours à la stratégie "*Min*". Les résultats d'une telle évaluation sont présentés sur le tableau suivant:

Tableau. 4.11: Résultats expérimentaux obtenus sur des images contenant des pixels peau de toutes les races sans utilisation de la stratégie "*Min*".

	Espace de couleur	VP	FP	Distance-point optimal	CDR	FAR	FRR
RLBN multinomiale	IQ	97.191041	16.106565	16.3497	**87.7499**	11.4355	0.8146
	YIQ	97.073864	47.838502	47.9279	65.1865	33.9649	0.8486
	CbCr	96.987358	16.247047	16.5240	87.5911	11.5352	0.8737
	YCbCr	**97.258338**	47.285827	47.3652	65.6324	33.5725	**0.7951**
	xyz	95.094543	**14.31027**	**15.1277**	88.4172	**10.1601**	1.4226
	rgb	96.230252	32.541186	32.5788	75.8028	23.1039	1.0933

Nous pouvons constater que le classifieur basé sur la RLBN multinomiale du tableau 4.10 est bien plus efficace que celui du tableau 4.11. En effet, hormis pour l'espace xyz, dans tous les espaces et sous-espace de couleurs utilisés le nombre de faux positifs augmente considérablement en omettant d'utiliser la stratégie "*Min*". La distance au point optimal connait dans ce cas une dégradation remarquable. Seul l'espace xyz semble ne pas être affecté par la non utilisation de cette stratégie. Cette stratégie semble donc être efficace pour réduire le taux de fausse détection "*FP*". Dans tous les cas, les meilleurs résultats sont ceux de la RLBN multinomiale en employant la stratégie "*Min*".

La figure 4.6 illustre la position des points correspondant à nos évaluations effectuées dans l'espace IQ présentées dans les tableaux 4.7, 4.8, 4.9 et 4.10. Nous avons utilisé les symboles suivants pour distinguer les différents classifieurs utilisés :

+ : classifieur basé sur la RLBN multinomiale.

* : classifieur basé sur la RLBN simple entrainé sur toutes les races.

o : classifieur basé sur la RLBN simple entrainé sur la race sur laquelle s'effectue l'évaluation en cours.

□ : classifieur optimal.

Nous avons utilisé une couleur pour chaque type d'évaluation comme suit : le bleu pour une évaluation sur toutes les races, le rouge pour évaluation sur la race blanche, le jaune pour une évaluation sur la race jaune et le noir pour une évaluation sur la race noire.

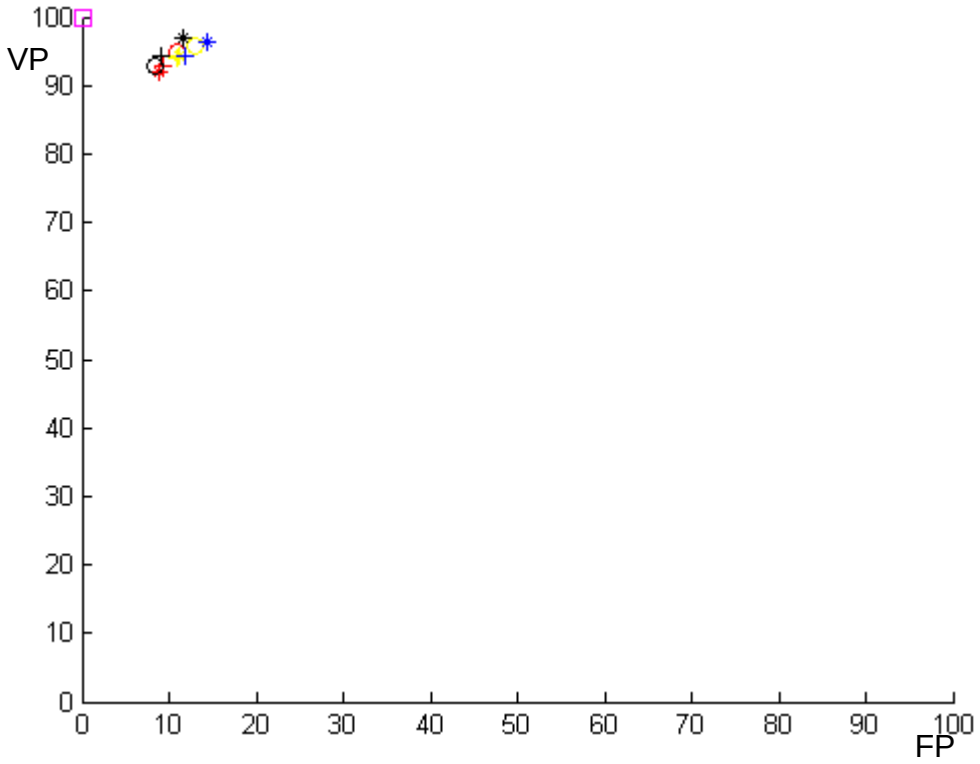
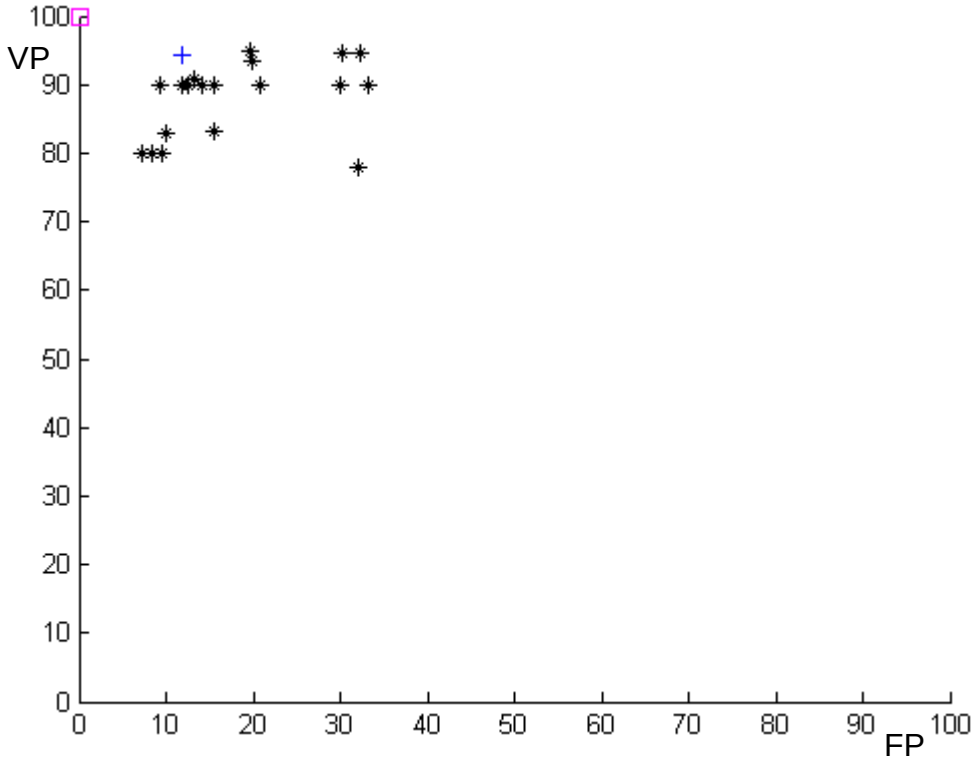


Figure. 4.6: Evaluations des différents classifieurs de type RLBN dans l'espace IQ.

En examinant la figure 4.6 de près, nous pouvons remarquer que les points correspondant au classifieur basé sur une RLBN multinomiale sont les plus proches du point optimal dans tous type d'évaluation effectué d'où sa supériorité par rapport aux autres classifieurs que nous avons expérimenté.

Nous compareront les résultats présentés sur tableau 4.10 à ceux du tableau 4.6. La plus petite distance au point optimal dans le tableau 4.6 est de 13.65. Cette performance est assurée par la méthode présentée dans [4]. Nous pouvons constater que ce score est parmi les meilleurs en comparaison à nos évaluations que nous avons effectuées sur la détection de toutes les races. Ils dépassent ceux de notre classifieur basé sur la RLBN simple. Néanmoins, nous notons une distance au point optimal de 13.0812 apporté par notre méthode basée sur la RLBN multinomiale dans l'espace IQ. Cette distance étant plus petite, nous pouvons constater la

supériorité de notre classifieur sur ceux du tableau 4.6. La figure 4.7 suivante illustre une comparaison des représentations des couples *(FP,VP)* obtenus comme résultats d'évaluation des méthodes citées sur le tableau 4.6 avec les résultats de notre approche basée sur la RLBN multinomiale dans l'espace IQ pour la détection de la peau appartenant à toutes les races.



+ : notre classifieur basé sur une RLBN multinomiale.

* : classifieurs du tableau 4.6.

□ : classifieur optimal.

Figure. 4.7: Comparaison des méthodes du tableau 4.6 avec notre classifieur basé sur la RLBN multinomiale.

Nous pouvons remarquer aisément que le point correspondant à notre classifieur est le plus proche du point (0,100) correspondant au classifieur optimal. Notre classifieur agit donc mieux que tous ceux basé sur les méthodes présentés dans le tableau 4.6.

Nous rappelons que les meilleurs résultats obtenus par la combinaison de réseaux de neurones dans [66] sont les suivants:

CDR = 82.61 % FAR = 14.66 % FRR = 2.73 %.

Ces résultats sont également surpassés clairement par ceux du classifieur basé sur la RLBN multinomiale dans les espaces IQ, CbCr et xyz ainsi que ceux du classifieur basé sur la RLBN simple dans les espaces YIQ, IQ, CbCr et YCbCr. Notre approche

fournit de meilleurs résultats sur tous les points à savoir: le taux de CDR est plus important dans le cas de notre approche et le taux de FAR ainsi que celui de FRR est moins important.

La figure 4.8 illustre trois exemples d'images de peau de différentes races et leurs masques binaires retournés par notre classifieur.

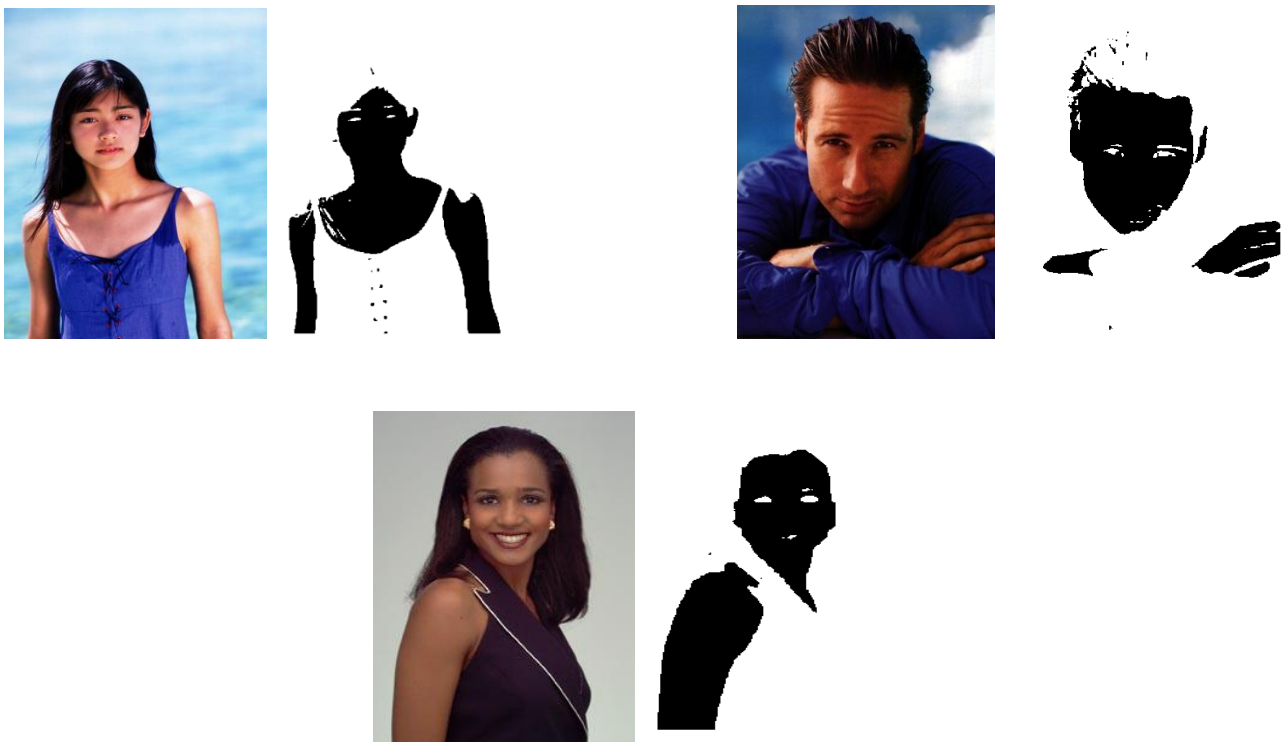


Figure. 4.8: Résultat de classification de pixels retournés par notre système.

4.6 Conclusion

Nous avons présenté dans ce chapitre un classifieur de peau et non peau dit RLBN basé sur une approche qui emploie à la fois les noyaux de Fisher, le théorème de Bayes et la régression logistique. Afin d'améliorer la performance de notre classifieur, nous nous sommes intéressé à traiter les pixels peau de trois races ethniques: la race jaune, la race noire et la race blanche séparément. A cet effet, nous avons développé un classifieur spécifique à chaque race. Puis, nous avons procéder à la construction d'un classifieur appelé "classifieur de synthèse" obtenu en combinant les classifieurs en appliquant une régression RLBN multinomiale. Nous avons utilisé à cet effet une combinaison de deux stratégies de fusion des classifieurs, à savoir la stratégie "Or logique" et la stratégie du "minimum". La stratégie "Or logique" permet d'améliorer le taux de bonne détection tandis que la stratégie du "minimum" permet de réduire le taux de fausse détection. Ce classifieur est donc destiné à la détection de tout type de peau de différentes races. Pour des fins de comparaison, nous avons entrainé et expérimenté un classifieur basé sur une RLBN simple sur des pixels peau correspondant à toutes les races. Les résultats ont

montré la supériorité du classifieur basé sur la RLBN multinomiale sur ce dernier. Nous avons employé la base d'image Compaq pour évaluer notre classifieur. Nous avons exploité plusieurs espaces et sous espaces de couleurs durant l'évaluation et les meilleurs résultats sont obtenus dans le sous espace IQ. Nous nous somme donc basé sur une évaluation effectuée sur ce sous espace pour comparer notre classifieur avec ceux fondés sur des méthodes proposées précédemment et qui ont été évalués sur la même base d'image que la notre. Après comparaison, nous avons constaté la supériorité de notre classifieur sur ces derniers.

CONCLUSION GENERALE

La modélisation de la peau humaine est un domaine de recherche qui consiste à localiser et détecter des personnes ou des parties du corps humains dans des images numérique ou vidéos. Trois types d'approches sont adoptées dans ce contexte: la détection basée sur l'extraction des traits caractéristiques, celle basée sur le mouvement et celle basée sur la couleur.
L'objectif de la modélisation de la peau est de construire des fonctions ou règles de décision permettant de discriminer les pixels de peau de ceux de non peau.
Celle-ci consiste en un travail préliminaire de base pour la construction de nombreuses applications: détection et reconnaissance de visages, filtrage de sites pornographiques, vidéo surveillance, etc.

Nous nous sommes intéressé à définir un modèle de détection de la peau sur des images fixes. De ce fait, Les approches basées sur le mouvement et l'extraction de traits caractéristiques ne correspondent pas à notre besoin. A cet effet, nous avons choisi d'adopter une méthode basée sur une approche couleur. La couleur fournit une information utile aux objets afin de faciliter leur localisation. L'objectif d'un système de détection de la peau est de pouvoir distinguer la couleur de la peau des autres couleurs de fond de l'image ou d'autres objets lui appartenant en élaborant une classification de type binaire: classe de peau et classe de non peau.

Plusieurs types de méthode de détection basés sur la couleur se distinguent. Les plus célèbres sont les méthodes paramétriques, non paramétriques et explicites. Les méthodes paramétriques se basent sur une estimation des densités des distributions colorimétriques des pixels traités. Les méthodes non paramétriques ne dépendent d'aucune distribution des pixels traités. Les méthodes explicites se basent sur une fixation empirique de règles de classification. La pertinence des méthodes paramétriques et leur qualité de précision dépendent largement de la forme de distribution des pixels traités ainsi que de l'espace de couleur de conversion choisi. Nous avons adopté une méthode non paramétrique pour la construction de notre système de détection de la peau. Il s'agit de la régression logistique bayésienne à noyau (*RLBN*). Il a été montré dans [131] que la régression logistique bayésienne à noyau donne de meilleurs résultats de classification que beaucoup d'autres méthodes basées sur les noyaux de Fisher comme les séparateurs à vaste marge (*SVM*), la méthode adaboost régularisé (AB_R), les machines à vecteurs de pertinence (*RVM*). Il s'agit d'une méthode basée sur un apprentissage supervisé consistant à estimer les paramètres du modèle de manière à séparer les pixels de peau de ceux de non peau de manière efficace en ayant recours au théorème de Bayes. La régression logistique est définie dans cette méthode comme modèle de prédiction permettant de construire une frontière linéaire entre les pixels de peau et ceux de non peau. Le noyau de Fisher est utilisé afin de projeter la représentation des pixels

traités de l'espace d'origine vers l'espace des caractéristiques. Celui-ci permet de maximiser la distance entre les moyennes des classes considérées et minimiser leurs variances intra-classes. La séparation des deux classes considérées est donc plus facile et efficace dans cet espace que dans l'espace d'origine. Nous avons utilisé un modèle *RLBN* pour la détection de la peau de chaque race ethnique. Nous avons considérés trois races: la race blanche, jaune et noire. La fusion des résultats de tous ces classifieurs constitue le résultat final de détection de tout type de peau basé sur une régression logistique bayésienne à noyaux multinomiale. Nous avons utilisé les règles "Or logique" et "Min" comme stratégies de fusion de ces classifieurs. La stratégie "Or logique" permet d'améliorer le taux de bonne détection car elle agit comme une fusion de tous les résultats des classifieurs considérés. Cependant, comme la structure de notre système est fondé sur la base de la fusion de plusieurs classifieur de type RLBN simple, le taux de fausse détection apporté par chacun de ces derniers additionné à ceux des autres contribue à la dégradation des performances de détection. La stratégie "Min" permet de réduire ce taux de fausse détection et d'améliorer les résultats. Avec un taux de bonne détection augmenté et un taux de fausse détection réduit, notre l'approche RLBN multinomiale agit mieux qu'une approche RLBN simple pour la détection de tout types de peau appartenant à différentes origines ethniques. L'évaluation effectuée sur la base d'images Compaq montre l'efficacité de notre approche en comparaison avec celles de l'état de l'art ayant effectué des évaluations sur cette même base d'images.

Perspectives

Le processus de détection de la peau correspond à une segmentation d'image. La segmentation consiste en une opération de traitement d'images qui a pour but de rassembler des pixels entre eux suivant des critères prédéfinis. Ces critères se basent généralement sur les caractéristiques dites de *bas niveau* de l'image. Ces caractéristiques sont la couleur, la texture et la forme. Notre modèle de détection est bien basé sur la couleur. Le processus de détection de notre système correspond à une segmentation de l'image traitée en région peau et non peau en s'appuyant sur l'information de couleur qui caractérisent celles-ci. Il est envisageable de tenir compte également des caractéristiques de texture de la peau et de la forme en plus de la couleur. L'association de telles caractéristiques avec celles de la couleur de la peau contribuerait à l'amélioration de notre système. Afin de mieux situer les performances de détection de notre classifieur par rapport à celle de l'état de l'art, il serait également intéressant d'approfondir l'évaluation sur d'autres bases d'images que Compaq. Nous pouvons aussi porter une amélioration notable à l'efficacité de notre système de détection en lui associant un algorithme de traitement de la variabilité de la luminance dans les images à traiter.

Notre modèle de peau est générique. Il pourrait tout à fait être utilisé comme une base pour la construction de diverses applications qui se basent sur l'identification ou la localisation de la peau comme la vidéo surveillance et le filtrage sur internet.

APPENDICE
REGRESSION LOGISTIQUE BEYESIENNE A NOYAU

1 Introduction

R.Ksantini et al. [131] ont proposé en 2008 une méthode de classification basée sur la régression logistique, les noyaux de Fisher et le théorème de Bayes dite "régression logistique bayésienne à noyaux" (*RLBN*). L'objectif de cette méthode est d'effectuer une projection linéaire de la représentation des pixels dans l'espace colorimétrique d'origine vers l'espace des caractéristiques de manière à pouvoir séparer les classes peau et non peau d'une manière plus efficace (pour plus de détails voir la section 3.9.1 du chapitre 3). L'application de la régression logistique dans ce nouvel espace permet de maximiser la distance entre les moyennes des classes considérées et minimiser leurs variances intra-classes. Nous aborderons dans cette annexe la méthode permettant l'estimation des paramètres du modèle de prédiction utilisé dans [131].

2 Estimation des paramètres du modèle de prédiction

Dans le cadre de l'apprentissage supervisé, nous supposons que l'échantillon d'apprentissage est représenté sous la forme : $\left\{ \tilde{X}, Y \right\}$ où : $\tilde{X} = \{X_i\}_{i=1}^{N}$ avec $X_i \in R^k$ ($k \geq 1$) $\forall i \in \{1,2\ldots..,N\}$, X_i étant la représentation du pixel i dans l'espace colorimétrique d'origine et $Y = \{t_i\}_{i=1}^{N}$ avec $t_i \in \{0,1\}$ une variable binaire indiquant la classe du pixel i. Si $t_i = 0$ alors le pixel i appartient à la classe non peau sinon il appartient à la classe peau.

Supposons que $\tilde{X}_1 = \{X_i\}_{i=1}^{N_1}$ et $\tilde{X}_2 = \{X_i\}_{i=N_1+1}^{N}$ deux classes différentes constituant l'espace initial de N vecteurs. En appliquant l'astuce du noyau, nous utilisons la fonction Φ pour représenter les classes $\tilde{X}_1$ et $\tilde{X}_2$ (classe peau et classe non peau) dans les espaces $F_1 = \{\Phi(X_i)\}_{i=1}^{N1}$ et $F_2 = \{\Phi(X_i)\}_{i=N1+1}^{N}$ respectivement.

avec : $\Phi(X_i) = (1, K(X_i, X_1), K(X_i, X_2),..., K(X_i, X_N)) \forall i \in \{1,2,..., N\}$

Notons Φ_1 Φ_2 deux vecteurs aléatoires appartenant respectivement à F_1 et F_2. On suppose que $\Phi_1 \sim g_1(\Phi_1)$ et $\Phi_2 \sim g_2(\Phi_2)$ où g_1 et g_2 représentent deux distributions gaussiennes différentes dont les moyennes et les matrices de covariance sont calculés empiriquement à partir de F_1 et F_2 respectivement.
Nous associons $t_1=0$ et $t_2=1$ à Φ_1 et Φ_2 respectivement.
Les paramètres (poids) associés à chaque fonction noyau sont définis d'une manière aléatoire. Soit le vecteur $w= (w_0, w_1,\ldots, w_N)$ qui dénote ces paramètres.

On suppose que $w \sim \pi(w \mid \beta) = \prod_{i=0}^{N} N(w_i \mid 0, \beta_i^{-1})$ avec $\beta = \{\beta_0, \beta_1, \beta_2, \ldots, \beta_N\}$ est le vecteur de N+1 hyperparamètres inconnus. Les paramètres β_i sont définis comme suit:

$P(\beta) = \prod_{i=0}^{N} Gamma(\beta_i \mid a, b)$ avec $Gamma(\beta_i \mid a, b) = \Gamma(a)^{-1} b^a \beta_i^{a-1} \exp(-b\beta_i)$ et

$\Gamma(a) = \int x^{a-1} \exp(-x) dx$ la fonction gamma.

En employant la règle de Bayes, la distribution à posteriori à travers tous les paramètres w et β est définie par:

$$P(w, \beta \mid t_1 = 0, t_2 = 1) = \frac{P(t_1 = 0, t_2 = 1 \mid w, \beta)\pi(w, \beta)P(\beta)}{P(t_1 = 0, t_2 = 1)} \quad (1)$$

avec:

$$P(t_1 = 0, t_2 = 1) = \int P(t_1 = 0, t_2 = 1 \mid w, \beta)\pi(w \mid \beta)P(\beta) dw d\beta \quad (2)$$

Il n'est pas possible de calculer la distribution à postériori des paramètres inconnus puisqu'on ne peut pas calculer l'intégrale $P(t_1 = 0, t_2 = 1)$.
Pour simplifier, on décompose $P(w, \beta \mid t_1 = 0, t_2 = 1)$ en deux distributions à postériori.

$P(w \mid t_1 = 0, t_2 = 1, \beta)$ et $P(\beta \mid t_1 = 0, t_2 = 1)$.

On approxime $P(w \mid t_1 = 0, t_2 = 1, \beta)$ avec une distribution gaussienne et on calcule les hyper paramètres optimaux qui maximisent l'approximation de $P(\beta \mid t_1 = 0, t_2 = 1)$.
La décomposition à travers tous les paramètres inconnus se fait donc de la manière suivante:

$P(w, \beta \mid t_1 = 0, t_2 = 1) = P(w \mid t_1 = 0, t_2 = 1, \beta)\ P(\beta \mid t_1 = 0, t_2 = 1)$

On définit la densité de probabilité à postériori de w selon l'équation suivante:

$$P(w \mid t_1 = 0, t_2 = 1, \beta) = \frac{P(t_1 = 0, t_2 = 1 \mid w)\pi(w \mid \beta)}{P(t_1 = 0, t_2 = 1 \mid \beta)} \quad (3)$$

avec:

$$P(t_1 = 0, t_2 = 1 \mid w) = \prod_{i=1}^{i=2} P(t_i = i-1 \mid w)$$

$$= \sum_{\Phi_i \in F_i} \frac{\prod_{i=1}^{i=2} P(t_i = i-1, \Phi_i, w)}{(P(w))^2} = \sum_{\Phi_i \in F_i} \prod_{i=1}^{2} \left[\frac{P(t_i = i-1, \Phi_i, w)}{P(w, \Phi_i) P(w)} \right] P(w, \Phi_i) \quad (4)$$

Avec : $P(w) = \int \pi(w \mid \beta) p(\beta) d\beta$

Dans le cadre de l'approche bayésienne, on considère généralement que l'espace des paramètres à estimer est indépendant de l'espace des observations. On suppose donc que w et Φ_i sont indépendants pour chaque i ∈ {1, 2}.
On considère alors que $P(w, \Phi_i) = g(\Phi_i) P(w)$ pour chaque i ∈ {1, 2}.

On obtient alors:

$$P(t_1 = 0, t_2 = 1 \mid w) = \sum_{\Phi_i \in F_i} \prod_{i=1}^{2} \left[\frac{P(t_i = i-1, \Phi_i, w)}{(P(w))^2 g_i(\Phi_i)} \right] g_i(\Phi_i) P(w)$$

$$= \sum_{\Phi_i \in F_i} \prod_{i=1}^{2} P(t_i = i-1 \mid \Phi_i, w) g_i(\Phi_i) \quad (5)$$

Dans le cadre de l'analyse discriminante, on suppose généralement que les distributions g_1 et g_2 sont indépendantes chacune de l'autre ainsi que du vecteur w. La probabilité conditionnelle $P(t_1 = 0, t_2 = 1 \mid w)$ peut alors s'écrire de la manière suivante:

$$P(t_1 = 0, t_2 = 1 \mid w) = \prod_{i=1}^{2} \left[\sum_{\Phi_i \in F_i} \left[P(t_i = i-1 \mid \Phi_i, w) g_i(\Phi_i) \right] \right] \quad (6)$$

avec:

$$P(t_i = i-1 \mid \Phi_i, w) = F((2i-3) w^T \Phi_i) \quad (7)$$

où :

$$F(x) = \frac{\exp(x)}{1 + \exp(x)} \quad (8)$$

$P(t_i = i-1 \mid \Phi_i, w)$ représente le modèle logistique de t_1 et t_2 étant donné $\Phi_1$et Φ_2.
Par conséquent, on a:

$$P(w \mid t_1 = 0, t_2 = 1, \beta) = \frac{\prod_{i=1}^{2}\left[\sum_{\Phi_i \in F_i}\left[P(t_i = i-1 \mid \Phi_i, w) g_i(\Phi_i)\right]\right]\pi(w \mid \beta)}{P(t_1 = 0, t_2 = 1 \mid \beta)} \quad (9)$$

avec:

$$P(t_1 = 0, t_2 = 1 \mid \beta) = \int \sum_{\Phi_i \in F_i}\left[\prod_{i=1}^{2} P(t_i = i-1 \mid \Phi_i, w) g_i(\Phi_i)\right]\pi(w \mid \beta) dw \quad (10)$$

La maximisation de la distribution $P(w, \beta \mid t_1 = 0, t_2 = 1)$ revient à maximiser $P(t_1 = 0, t_2 = 1 \mid w)$. Cela permet donc d'effectuer la projection linéaire de g_1 et g_2 la plus discriminante.
Le calcul de la distribution à posteriori $P(w \mid t_1 = 0, t_2 = 1, \beta)$ est intraitable puisqu'elle est basée sur le traitement d'une intégrale multidimensionnelle. Cette dernière ne peut être maximisée analytiquement. Cependant, on peut l'estimer par une approximation variationnelle à posteriori avec une forme gaussienne dont le calcul de la moyenne et de la matrice de covariance est faisable. D'après [131], une approximation est faite comme suit:

$$\sum_{\Phi_i \in F_i}\left[\prod_{i=1}^{2} P(t_i = i-1 \mid \Phi_i, w) g_i(\Phi_i)\right] \geq \left[\prod_{i=1}^{2} F(\varepsilon_i)\right] \exp\left[\sum_{i=1}^{2}\left[\frac{E_{gi}[H_i] - \varepsilon_i}{2}\right] - \sum_{i=1}^{2}\left[\varphi(\varepsilon_i)(E_{gi}[H_i^2] - \varepsilon_i^2)\right]\right]$$

$$= P(t_1 = 0, t_2 = 1 \mid w, \{\varepsilon_i\}_{i=1}^{2}) \quad (11)$$

avec $H_i = (2i-3) w^T \Phi_i \;\; \forall i \in \{1,2\}$ et F(x) la régression logistique définie précédemment avec $\varepsilon_i > 0$ est le paramètre variationnel,

$$\varphi(\varepsilon_i) = \frac{\tanh(\frac{\varepsilon_i}{2})}{4\varepsilon_i} \quad (12)$$

où:

$$\tanh(\frac{\varepsilon_i}{2}) = \frac{\exp(\frac{\varepsilon_i}{2}) - \exp(-\frac{\varepsilon_i}{2})}{\exp(\frac{\varepsilon_i}{2}) + \exp(-\frac{\varepsilon_i}{2})} \quad \forall i \in \{1,2\} \tag{13}$$

E_{g1} et E_{g2}: représentent les estimations discrètes par rapport aux distributions gaussiennes g_1 et g_2 respectivement. On a donc:

$$\sum_{\Phi_i \in F_i} \left[\prod_{i=1}^{2} P(t_i = i-1 \mid \Phi_i, w) g_i(\Phi_i) \right] \pi(w \mid \beta) \geq P(t_1 = 0, t_2 = 1 \mid w, \{\varepsilon_i\}_{i=1}^{2}) \pi(w \mid \beta) \tag{14}$$

Ainsi, l'approximation à posteriori dénotée par $P(w \mid t_1 = 0,\ t_2 = 1, \{\varepsilon_i\}_{i=1}^{2}, \beta)$ est effectué en trouvant la bande minimale présentée dans l'équation (A.14). L'approximation variationnelle à posteriori est de type gaussienne. Elle est caractérisée par sa moyenne à posteriori notée μ_{post} et sa matrice de covariance notée $\sum_{post}$ calculées comme suit:

$$(\sum\nolimits_{post})^{-1} = A^{-1} + 2\sum_{i=1}^{2} \left[\varphi(\varepsilon_i) E_{gi} \left[\Phi_i \Phi_i^T \right] \right] \tag{15}$$

$$\mu_{post} = \sum\nolimits_{post} \left[\sum_{i=1}^{2} \left[(i - \frac{3}{2}) E_{gi} \left[\Phi_i \right] \right] \right] \tag{16}$$

avec :

$A = diag(\beta_0^{-1}, \beta_1^{-1}, ..., \beta_N^{-1})$

Selon l'équation (1), $\sum_{post}$ dépend des paramètres variationnels $\{\varepsilon_i\}_{i=1}^{2}$ et $\{\beta_i\}_{i=1}^{N}$ qui assure l'estimation d'une borne inférieure dans l'équation (14) et maximise ainsi l'approximation de $P(\beta \mid t_1 = 0, t_2 = 1)$. Les paramètres variationnels sont calculés comme suit:

$$\varepsilon_i^2 = E_{gi} \left[\Phi_i^T \sum\nolimits_{post} \Phi_i \right] + \mu_{post}^T \left[E_{gi} \left[\Phi_i \Phi_i^T \right] \right] \mu_{post} \forall i \in \{1,2\} \tag{17}$$

$$\beta_j = \frac{1 + 2a}{\sum_{post,jj} + \mu_{post,j}^2 + 2b} \quad \forall j \in \{0,1,..., N\} \tag{18}$$

3 Algorithme de calcul

L'algorithme d'estimation des poids w_i s'opère en deux phases. La première phase concerne l'initialisation, tandis que la seconde est itérative et permet le calcul de $\sum_{post}$

et μ_{post} comme indiqué dans les équations (A.15) et (A.16) tout en employant les équations (A.17) et (A.18) afin d'estimer les paramètres variationnels et les hyperparamètres à chaque itération. Les valeurs du vecteur μ_{post} représentent l'estimation recherché des poids $\{w_i\}_{i=1}^{N}$.

L'algorithme d'estimation se déroule comme suit :

Initialisation

1. Calcul des moyennes et matrices de covariances des gaussiennes g_1 et g_2 empiriquement à partir de F_1 et F_2 respectivement.
2. Initialisation des paramètres variationnels $\{\varepsilon_i\}_{i=1}^{2}$ et des hyperparamètres $\{\beta_i\}_{i=1}^{2}$.

Calcul de $\sum_{post}$ et μ_{post} :

1. **Do**

$$(\sum\nolimits_{post}^{new})^{-1} \leftarrow (A^{-1} + 2\sum_{i=1}^{2}\left[-\varphi(\varepsilon_i^{old})E_{gi}\left[\Phi_i\Phi_i^T\right]\right])$$

$$\mu_{post}^{new} \leftarrow \sum\nolimits_{post}^{new}\left[\sum_{i=1}^{2}\left[(i-\frac{3}{2})E_{gi}\left[\Phi_i\right]\right]\right]$$

For each $i \in \{1,2\}$ **do**

$$(\varepsilon_i^{new})^2 \leftarrow E_{gi}\left[\Phi_i^T \sum\nolimits_{post}^{new} \Phi_i\right] + (\mu_{post}^{new})^T\left[E_{gi}\left[\Phi_i\Phi_i^T\right]\right]\mu_{post}^{new}$$

End for

For each $j \in \{0,1,..., N\}$ **do**

$$\beta_j^{new} \leftarrow \frac{1+2a}{\sum_{post,jj}^{new} + (\mu_{post,j}^{new})^2 + 2b}$$

End For

While $\left(\left|\mu_{post}^{old} - \mu_{post}^{new}\right| > threshold\right)$

Return μ_{post}^{new}

2. Affecter les valeurs des composantes de μ_{post} aux poids $\{w_i\}_{i=0}^{N}$.

4 Rôle de la parcimonie dans le calcul algorithmique

A fur et à mesure que l'optimisation de l'estimation des valeurs des hyperparamètres β_i progresse, l'intervalle d'appartenance de ces derniers s'élargit et peut s'étendre à l'infini. En effet, pour a=b=0, beaucoup d'hyperparamètres convergeront à l'infini et ces derniers ne peuvent contribuer dans le calcul. La

parcimonie provient de la troncature du coût. Elle permet de simplifier l'estimation en ne considérant que les hyperparamètres qui contribuent réellement dans le calcul. Si on considère un hyperparamètre $\beta_k \to \infty$ pour lequel on choisit k=1 (le premier hyperparamètre) pour des raisons de simplification, alors la matrice de covariance aura la forme suivante:

$$\sum_{post} \to \begin{pmatrix} 0 & 0 \\ 0 & (A_{-k}^{-1} + M_{-k})^{-1} \end{pmatrix} \tag{19}$$

Cette forme est obtenue en suspendant la participation de β_k dans le calcul.

avec :

$$M = 2\sum_{i=1}^{2}\left[-\varphi(\varepsilon_i) E_{gi}\left[\Phi_i \Phi_i^T\right]\right] \tag{20}$$

L'indice "-k" utilisé correspond à la matrice à laquelle il se rapporte avec suppression des $k^{ièmes}$ ligne et colonne. Le terme $(A_{-k}^{-1} + M_{-k})^{-1}$ correspond à la matrice de covariance à posteriori calculée sans la fonction de base k. On note donc une réduction du nombre de fonctions de bases et donc de la complexité algorithmique.

REFERENCES

1. K. Aas, "Detection and recognition faces in video sequences", Norsk regnesentral, 1998.

2. S.G. Ababsa, "Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D". Thèse de doctorat en informatique à l'université Evry Val d'Essonne, 03 Octobre 2008.

3. M. Abdel-Mottaleb and A. Elgammal, "Face detection in complex environments from color lmages". In: Proceedings of the International Conference on Image Processing (ICIP). 622–626,1999.

4. M. Abdullah-Al-Wadud, M. Shoyaib and O. Chae, "A Skin Detection Approach Based on Color Distance Map". Hindawi Publishing Corporation EURASIP Journal on Advances in Signal Processing Article ID 814283, 10 pages, 16 December 2008.

5. A. Agresti, "Categorical Data Analysis". Chapter 4, "Models for Binary Response Variables". pages 79-129, Wiley, 1990.

6. A. Agresti, "An Introduction to Categorical Data Analysis". JohnWiley and Sons Inc., 1996.

7. J. Ahlberg, "Extraction and Coding of Face Model Paramaters". Licentiate Thesis No. 747, Department of Electrical Engineering, Linlkoping University, Sweden, March 1999.

8. S. Ahmad, "A usable real-time 3d hand tracker". Conference Record of the Asilomar. Conference on Signals, Systems and Computers, pp. 1257-1261, 1994.

9. M. Aizerman, E. Braverman, and L. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning". Dans Automation and Remote Control, vol. 25, 1964, p. 821-837.

10. M. Alan and McIvor, "Background subtraction techniques". IVCNZ, 2000.

11. S. Attaway, "Matlab: A Practical Introduction to Programming and Problem

Solving", College of Engineering, Boston University Boston, MA, Elsevier, 2009.

12. M.M. Aznaveh, H. Mirzaei, E. Roshan, and M. Saraee, "A new and improved skin detection method using RGB vector space". In Proceedings of IEEE International Multi- Conference on Systems, Signals and Devices, July 2008, pp.1–5.

13. F.R. Bach, G.R.G. Lanckriet and M. I. Jordan, "Multiple kernel learning, conic duality, and the smo algorithm". In ICML'04: Proceedings of the twenty first international conference on Machine learning, page 6, New York, NY, USA, ACM Press, 2004.

14. M. Bardos, "Analyse discriminante - Application au risque et scoring financier, Chapitre 3, Discrimination logistique". pages 61-79, Dunod, 2001.

15. F. Bérard, "Vision par ordinateur pour l'interaction homme-machine fortement couplée". Thèse de doctorat à l'université Joseph Fourier, Grenoble, France, Janvier 2000.

16. K.G. Bernhard Fink and P.J. Matts, "Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces". Evolution and Human Behavior 27(6) 433–442, 2006.

17. K.K. Bhoyar and O.G. Kakde, "Skin Color Detection Model Using Neural Networks and its Performance Evaluation". Journal of Computer Science 6 (9): 963-968, ISSN 1549-3636, 2010.

18. C. Bishop, "Neural Networks for Pattern Recognition". Clarendon Press, Oxford, 1995.

19. C. Bishop, "Pattern Recognition and Machine Learning", Springer, 2006.

20. S. Birchfield, "Elliptical head tracking using intensity gradients and color histograms". In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Santa Barbara, CA, pp. 232-237, 1998.

21. B.E. Boser, I. M. Guyon and V.N. Vapnik, "A training algorithm for optimal margin classifiers", Dans Proceedings of the 5th ACM Workshop on Computational Learning Theory (COLT), pages 144–152, 1992.

22. G. Bouchard, "Les modèles génératifs en classification supervisée et applications à la catégorisation d'images et à la fiabilité industrielle", thèse de doctorat à l'université Joseph Fourier-Grenoble1, 2005.

23. N. Bouguila and D. Ziou, "Dirichlet-based probability model applied to human skin detection". IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004.

24. N. Bouguila and D. Ziou, "Unsupervised Learning of a Finite Discrete Mixture:

Applications to Texture Modeling and Image Databases Summarization". IEEE transaction on pattern analysis and machine intelligence, VOL. 29, No. 10, October 2007.

25. N. Bouguila, D. Ziou and E. Monga, "Practical bayesian estimation of a finite beta mixture through gibbs sampling and its applications". Statistics and Computing, vol. 16, pp. 215–225, 2006.

26. J. Brand and J.S. Mason. "A comparative assessment of three approaches to pixel-level human skin-detection". In Proceedings 15th International Conference on Pattern Recognition, vol. 1, Barcelona, Spain, pp. 1056-1059, September 2000.

27. D. Brown, I. Craw, and J. Lewthwaite, "A som based approach to skin detection with application in real time systems". In Proc. of the British Machine Vision Conference, volume 2, pp. 491-500, 2001.

28. R. Brunelli and T. Poggio. "Face recognition: features versus templates". IEEE-T-PAMI, Vol. 15, No. 10, pp.1042-1052, Octobre 1993.

29. Budget fédéral 2002. http://www.radio-canada.ca/nouvelles/budget/federal_2002/ bref.html.

30. Budget fédéral 2003. http://www.fin.gc.ca

31. M.C. Burl and P. Perona, "Recognition of planar object classes". Comp. Vision and Pattern Recognition, San Fransisco, Juin 1996.

32. T.S. Caetano, S. D. Olabarriaga, and D. A. C. Barone, "Do mixture models in chromaticity space improve skin detection?". Pattern Recognition, 36: 3019_3021, 2003.

33. J. Cai, A. Goshtasby, and C. Yu, "Detecting human faces in color images". Int'l Workshop on Multi-Media Database Management Systems, pp.124-131, 1998.

34. F. H. Cardoso and H. M. Gomes, "A Probabilistic Approach to Skin Detection", In Anais do Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens 2007.

35. G.C. Cawley et N.L.C. Talbot, "Sparse Bayesian Kernel Logistic Regression", ESANN'2004 proceedings - European Symposium on Artificial Neural Networks Bruges (Belgium), 28-30 Avril 2004.

36. D. Chai, K.N. Ngan, "Face segmentation using skin-color map in videophone applications". IEEE Trans. Circuits Syst. Video Technol.9 (4) 1999.

37. D. Chai and A. Bouzerdoum, "A bayesian approach to skin color classification in YCbCr color space". In Proc. IEEE Region Ten Conference (TENCON'2000), Vol. 2, pp. 421-424, 2000.

38. D. Chai, S.L. Phung and A. Bouzerdoum, "A Bayesian skin/non-skin color classifier using non-parametric density estimation". In Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on, May 2003, vol. 2, pp. II–464–II–467 vol.2.

39. Y. Chahir and A. Elmoataz (2006), "Skin-color Detection using Fuzzy Clustering". vISCCSP 2006.

40. Q. Chen, H. Wu and M. Yachida. "Face detection by fuzzy pattern matching". Proc. 5th Int. Conf. on Computer Vision, pp. 591-596, Cambridge, 1995.

41. H.D. Cheng, X.H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects". Pattern Recognition, 34(12), pp. 2259-2281, December 2001.

42. B. Cheng, "Automatic Vessel and telegiectases analysis in dermoscopy skin lesion images". Thesis in computer engineering Presented to the Faculty of the Graduate School of the Missouri University of Science and Technology, 2009.

43. S.L Chiu, "Fuzzy Model Identification Based on Cluster Estimation, Journal of Intelligent and Fuzzy Systems", 2, 267-278, 1994.

44. S.L Chiu, "Extracting Fuzzy Rules from Data for Function Approximation and Pattern Classification". In Dubois, D., Prade, H., and Yager, R. (Eds.), Fuzzy Information Engineering: A Guided Tour of Applications. John Wiley & Sons, Inc., New York, USA, 1997.

45. C.C. Chiang, W.N. Tai, M. T. Yang, Y. T. Huang and C. J. Huang, "A novel method for detecting lips, eyes and faces in real time". Real-Time Imaging, 9, 2003.

46. T.Y. Chin, "Fuzzy Skin detection". A project for the award of the degree of Master of Science (Computer Science) in Faculty of Computer Science and Information System Universiti Teknologi Malaysia, October 2008.

47. T. Choi, M.J. Schervish, K.A. Schmitt et M.J. Small, "A Bayesian approach to logistic regression model with incomplete information", PMID: 17764482 PubMed - indexed for MEDLINE, 30 Avril 2006.

48. O. Chomat, "Caractérisation d'éléments d'activités par la statistique conjointe de champs réceptifs". Thèse de doctorat, Institut National Polytechnique de Grenoble, Juin 2000.

49. D. Collett, "Modelling binary data", London, Chapman & Hall, 369P, 1991.

50. R. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade, "Algorithms for cooperative multisensor surveillance". In Proceedings of the IEEE, 89(10): 1456_1477, 2001.

51. M. Collobert, R. Feraud, G.L. Tourneur, D. Bernier, J.E. Viallet, Y. Mahieux, and D. Collobert, "A system for locating and tracking individual speakers".

International Conference on Automatic Face and Gesture Recognition, pp. 283-288, October 1996.

52. E. Côme, "Apprentissage de modèles génératifs pour le diagnostic de systèmes complexes avec labellisation douce et contraintes spatiales", thèse de doctorat à l'université de technologie de Compiègne Heudiasyc, Inrets-LTN, 16 janvier 2009.

53. Commission Internationale de l'Eclairage. Colorimetry. Rapport technique 15.2, Bureau central de la CIE, Vienna, 1986.

54. T.F. Cootes, C.J. Taylor, D.H. Cooper and J. Graham, "Active shape models-their training and application", Comp. Vision and Image Understanding, Vol. 61, No. 1, pp. 38-59, 1995.

55. D.R. Cox, "The analysis of binary data", London, Methuen, 1970.

56. D.R. Cox, E.J.Snell, "Analysis of binary data", London, Chapman & Hall, 236P, 1989.

57. M. Craven and J.W. Shavlik, "Extracting tree-structured representations of trained networks". Advances in Neural Information Processing Systems, pp. 24-30, 1996.

58. I. Craw, D. Tock and A. Bennett, "Finding face features". Proc. 2nd European Conference on computer Vision, pp. 92-96, 1992.

59. N. Cristianini and J. Shawe-Taylor, "An Introduction to Support Vector Machines and other kernel-based learning methods". Cambridge University Press, Cambridge, UK, 2000.

60. R. Cucchiara, C. Grana, M. Piccardi, A. Pratti and S. Sirotti, "Improving shadow suppression in moving object detection with hsv color information". Intelligent Transportation Systems, pp. 334– 339. IEEE, 2001.

61. H. Deng, E.N. Mortensen, L. Shapiro, and T.G. Dietterich, "Reinforcement matching using region context". In Proc. "beyond patches" CVPR Workshop, page 11, 2006.

62. L. Denoyer, "Apprentissage et inférence statistique dans les bases de documents structurés: Application aux corpus de documents textuels". Thèse de doctorat à l'université Paris 6,15 Décembre 2004.

63. T.G. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes", Journal of Artificial Intelligence Research, 2: 263–286, 1995.

64. A. Diplaros, T. Gevers and N. Vlassis, "Skin Detection Using The EM Algorithm with Spatial Constraints". IEEE International Conference, Vol 4, 3071 – 3075, October 2004.

65. P. Domingos and M. J. Pazzani. Beyond independence: Conditions for the optimality of the simple Bayesian classifier. In International Conference on Machine Learning, pages 105–112, 1996.

66. C.A. Doukim, Jamal Ahmad Dargham, Ali Chekima et Sigeru Omatu, "Cmbining neural networks for skin detection". Signal & Image Processing: An International Journal(SIPIJ) Vol.1, No.2, December 2010.

67. R.O. Duda and P.E. Hart, "Pattern classification and scene analysis". John Wiley, 1973.

68. R.O. Duda, P.E. Hart and D.G. Stork. Pattern Classification (2nd Edition). Wiley, 2000.

69. F. Duyme et J.J. Claustriaux, "La régression logistique binaire", Notes de statistique et d'informatique, Faculté universitaire des sciences agronomiques. Unité de statistique, informatique et mathématique appliquées, Gembloux (Belgique), 2006.

70. ECU images basis, "http://www.some.ecu. au/~sphung".

71. A. Elgammal, D. Harwood and L. Davis, "Non-parametric model for background subtraction". European Conference on Computer Vision, pp. 751–767, 2000.

72. X. Fan, L. Wang, "Comparing linear discriminant function with logistic regression for the two-group classification problem", J.Experim.Edu.67, 265-286, 1999.

73. O.D. Faugeras, "Digital color image processing and psychophysics within the framework of a human visual model". Ph.d. dissertation, University of Utah, Juin 1976.

74. R.A. Fisher, "The use of multiple measurements in taxonomic problems". Annals of Eugenics, 7: 179–188, 1936.

75. R.A. Fisher, "The use of multiple measurements in taxonomic problems". Annals of Eugenics 7, pp. 179-188, reprinted in Contributions to Mathematical statistics, John Wiley, New-York, 1950.

76. S. Fkihi, M. Daoudi, D. Aboutajdine, "Skin and non-skin probability approximation based on discriminative tree distribution". Image Processing (ICIP), 16th IEEE International Conference, 7-10 Nov. 2009. Pp. 2377 – 2380, Morocco, November 2009.

77. M.M. Fleck, D.A. Forsyth and C. Bregler, "Finding naked people". European Conference on Computer Vision, Springer-Verlag, Berlin, Germany, pp. 592-602, 1996.

78. D.A. Forsyth and M.M. Fleck, "Identifying nude pictures". IEEE Workshop on

the Applications of Computer Vision, pp. 103-108, 1996.

79. J.H. Friedman, "Another approach to polychotomous classification", Rapport technique, Department of Statistics, Stanford University, 1996. Disponible à l'adresse http://www-stat.stanford.edu/ jhf/ftp/poly.ps.Z.

80. L.M. Fu, "Neural networks in computer intelligence", McGraw-Hill, 1994.

81. C. Garbay, "Modélisation de la couleur dans le cadre de l'analyse d'images et de son application à la cytologie automatique". Thèse de doctorat, Institut National Polytechnique de Grenoble, Décembre 1979.

82. C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis". IEEE Trans. Multimedia 1 (3) 264–277, 1999.

83. F. Gasparini and R. Schettini, "Skin segmentation using multiple thresholding". In Internet Imaging VII, vol. 6061 of Proceedings of SPIE, pp. 1–8, San Jose, Calif, USA, January 2006.

84. V. Girondel, "Détection de peau, suivi de tête et de mains pour des applications multimédia". DEA Signal Image Parole Télécom, Juillet 2002.

85. M.R. Girgis, T.M. Mahmoud and T. Abd-el-hafeez, "An Approach to Image Extraction and Accurate Skin Detection from Web Pages". World Academy of Science Engineering and technology, 2007.

86. J.A.D. Gobineau,"L'Essai sur l'inégalité des races humaines". Diplomate et écrivain français, 1816-1882, créateur d'un système de pensée raciste 1854.

87. G. Gomez, "On selecting color components for skin detection". In Proc of the Int'l Conf. on Pattern Recognition, volume 2, pages 961_964, 2000.

88. G. Gomez, "On selecting colour components for skin detection". In Proceedings. 16th International Conference on Pattern Recognition, vol. 2, pp. 961-964, 2002.

89. R.C Gonzalez and R.E Woods, "Digital Image Processing", Prentice Hall, 2002.

90. G. Govaert, éditeur. Analyse des données. Hermès, 2003.

91. V. Govindaraju, D.B. Sher, R.K. Srihari and S.N. Srihari, "locating human faces in newspaper photographs". Proc. Conf. On Comp. Vision and Pattgern Recognition, pp. 549-554, 1989.

92. H.P. Graf, T. Chen, E. Petajan, and E. Cosatto, "Locating faces and facial parts". In Proc First Int'l Workshop Automatic Face and Resture Recognition, pages 41_46, 1995.

93. H.P. Graf, E. Cosatto, D. Gibbon, M. Kocheisen and E. Patajan, "Multimodal

system for locating heads and faces". Proc.2nd Int. Conf. On Automatic Face and Gesture Recognition, pp. 88-93, Vermont, 1996.

94. P. Green and B. Silverman, "Nonparametric regression and generalized linear models", Monographs on statistics and probability, Chapman & Hall, 1994.

95. H. Greenspan, J. Goldberger and I. Eshet, "Mixture model for face-color modeling and segmentation". Pattern Recognition Letters, 22: 1525_1536, 2001.

96. M. Hammami, "Modèle de peau et application à la classification d'images et au filtrage des sites Web". Thèse de doctorat en informatique à l'école nationale de Lyon, 2005.

97. K. Hammouda and F. Karray, "A Comparative Study of Data Clustering Techniques". SYDE 625: Tools of Intelligent Systems Design, Course project, http://pami.uwaterloo.ca/~hammouda/publications.php (July 18, 2008), 2000.

98. M.B. Hamid and Y.B. Jemma (2005), "Fuzzy Classification, Image Segmentation and Shape Analysis for Human Face Detection". ICSP2006 Proceedings.

99. T. Hastie, T. Tibshirani and J. Friedman, "The Elements of Statistical Learning, Data Mining, Inference and Prediction". Statistics, Springer, 2006.

100. T. Hastie and R. Tibshirani, "Classification by pairwise coupling", In Michael I. JORDAN, Michael J. KEARNS et Sara A. SOLLA, réds., *Advances in Neural Information Processing Systems*, volume 10, The MIT Press, 1998. Disponible à l'adresse citeseer.ist.psu.edu/hastie98classification.html.

101. R. Hérault et Yves Grandvale, "Régression Logistique Parcimonieuse". Manuscrit auteur, publié dans "9ème Conférence d'Apprentissage, Plate-forme AFIA", Grenoble, France, 2007.

102. T. Horprasert, D. Harwood and L. Davis, "A robust background subtraction and shadow detection". In Proceedings of the ACCV, 2000.

103. D.W. Hosmer, S. Lemeshow, "Applied logistic regression". New York, Wiley, 307 p, 1989.

104. D.W. Hosmer, S. Lemeshow, "Applied Logistic Regression", Second Edition, Wiley, 2000.

105. H. Hotelling, "Analysis of a complex of statistical variables into principal components", Journal of Educational Psychology, 24: 417–441,498–520, 1933.

106. J.C.V. Houwelingen, S. Le Cessie, "Logistic regression, a review". Stat.Neerl. 42,215-252., 1988.

107. R.L. Hsu, M. Abdel-Mottaleb and A.K. Jain, "Face detection in color images".

IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5), pp. 696-706, May 2002.

108. M. Hunke and A. Waibel, "Face locating and tracking for humain computer interaction". IEEE Computer, pp. 1277-1281, November 1994.

109. N. Idrissi, "La navigation dans les bases d'images: prise en compte des attributs de texture". Thèse de doctorat en informatique à l'université de Nantes, 18 Octobre 2008.

110. Imagis technologies inc. http://www.72technologies.com

111. M. Ivana Mikiæ, Pamela C. Cosman, Greg T. Kogut and Mohan M. Trivedi, "Moving shadow and object detection in scenes". In International Conference on Pattern Recognition (ICPR), pp. 321–324, 2000.

112. A. Jacquin and A. Eleftheriadis, "Automatic location tracking of faces and facial features in video signal". Int. Work. On Automatic Face and Gesture Recognition, pp.
142-147, Zurich, 1995.

113. T. Jebara, "Machine Learning: Discriminative And Generative". Kluwer Academic Publishers, 2004.

114. B. Jedynak, H. Zheng, M. Daoudi and D. Barret, "Maximum entropy models for skin detection". Tech. Rep. XIII, Universite des Sciences et Technologies de Lille, France, 2002.

115. M.J. Jones and J.M. Rehg, "Statistical color models with application to skin detection". Technical Report Cambridge Research Laboratory, CRL 98/11, Compaq, 1998.

116. M.J. Jones, and J.M. Rehg. "Statistical color models with application to skin detection". In Proc. of the CVPR '99, vol. 1, pp. 274–280, 1999.

117. M.J. Jones and J.M. Rehg. "Statistical color models with application to skin detection". International Journal of Computer Vision, 46(1), pp. 81-96, January 2002.

118. L. Jordao, M. Perrone, J.P. Coasteira and J. Santose-Victor, "Active Face and Feature Tracking". In Proceedings of the 10th International Conference on Image Analysis and Processing. 572 -577, 1999.

119. P. Kakumanu, S. Makrogiannis, N. Bourbakis, "A survey of skin-color modeling and detection methods". Pattern Recognition 40- 1106 – 1122, 2007.

120. T. Kanade, "Picture processing system by computer complex and recognition of humain faces". Phd thesis, Departement of Information Science, Kyoto University, Novembre 1973.

M. Kantardzic, "Data Mining: Concepts, Models, Methods, and Algorithms".

121. John Wiley & Sons, 2003 (343 pages).

122. J. Karlekar and U.B. Desai, 1999, "Finding faces in color images using wavelet transform". Proceeding of the International Conference on Image Analysis and Processing, Sept. 27-29, IEEE Computer Society, Washington, DC., USA., pp: 1085-1088. DOI: 10.1109/ICIAP.1999.797744.

123. V. Kettnaker and R. Zabih, "Bayesian multi-camera surveillance". In Proc of IEEE Conf. on Computer Vision and Pattern Recognition, pages 253_259, 1999.

124. R. Kjeldsen and J. kinder, "Finding skin in color images". Proc. 2nd Int. Conf. on Automatic Face and Gesture Recognition, pp. 312-318, Vermont, 1996.

125. G.J. Klinker, S.A. Shafer and T. Kanade, "A physical approach to color image understanding". International Journal of Computer Vision, Vol. 4, pp 7-38, 1990.

126. T. Kohonen, "Self-Organizing Maps". Springer Verlag, 1995.

127. T. Kohonen, "Self-Organizing and Associative Memory". Springer Verlag, 2ème edition, New York, 1984.

128. S.G. Kong, J. Heo, B.R. Abidi, J.Paik, and M.A. Abidi, "Recent advances in visual and infrared face recognition: a review". Computer Vision and Image Understanding, 97: 103_135, 2005.

129. J. Kovac, P. Peer, and F. Solina, "2D versus 3D colour space face detection". In Proceedings of the 4th EURASIP Conference on Video/Image Processing and Multimedia Communications, vol.2, pp. 449–454, Zagreb, Croatia, July 2003.

130. J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer, "Multi-camera multi-person tracking for easyliving". In Proc of IEEE Intl. Workshop on Visual Surveillance, pages 3_10, 2000.

131. R. Ksantini, D. Ziou and B. Colin and F. Dubeau "A Novel Bayesian Kernel Logistic Discriminant Model Based on a Variational Method". AAAI'08 Proceedings of the 23rd national conference on Artificial intelligence - Volume 3, 2008.

132. G. Kukharev and A. Novosielski, "Visitor identification elaborating real time face recognition system". In Proc. 12th Winter School on Computer Graphics (WSCG), Plzen, Czech Republic, pp. 157 – 164, Feb. 2004.

133. R. Kumar and A.Bindu, "An Efficient Skin Illumination Compensation Model For Efficient Face Detection". Proceeding of IEEE 32nd annual conference on Industrial Electronics, pp 3444-3449, November 2006.

134. P. Langley and S. Sage, "Tractable average-case analysis of naive bayesian classifiers". In M. Kaufmann, editor, Sixteenth International Conference on

Machine Learning, pages 220–228, Bled, Slovenia, 1999.

135. [135] H. Larochelle, "Étude de techniques d'apprentissage non-supervisé pour l'amélioration de l'entraînement supervisé de modèles connexionnistes", thèse de doctorat en informatique à l'université de Montréal, Décembre 2008.

136. J.Y. Lee and S.I. Yoo, "An elliptical boundary model for skin color detection". In Proc. International Conference on Imaging Science, Systems and Technology, Las Vegas, USA, June 2002.

137. A. Lemieux, "Système d'identification de personnes par vision numérique". Mémoire présenté à la Faculté des études supérieures de l'université Laval pour l'obtention du grade de maître en sciences (M.Sc.), Décembre 2003.

138. T. Leung, M. Burl and P. Perona, "Finding faces in cluttered scenes using labelled random graph matching". Proc. 5th Int. Conf. on Comp. Vision, pp. 637-644, Cambridge, 1995.

139. C. Lin, C.H. Su, H.S. Huang, and K.C. Fan, "Colour Image Segmentation Using Relative Values of RGB in Various Illumination Circumstances". International journal of computers. Issue 2, Volume 5, 2011.

140. S.P. Lloyd, "Least squares quantization in pcm". Information Theory, IEEE Transactions on, 28(2): 129–137, 1982.

141. D.G. Lowe, "Robust Model-based Motion Tracking Through the Integration of Search and Estimation". International Journal of Computer Vision, 8: 2, pp. 113-122, 1992.

142. X. Lu and A.K. Jain, "Ethnicity Identification from Face Images". In: Proceedings of SPIE International Symposium on Defense and Security: Biometric technology for human identification, 2004.

143. I. Ludmila Kuncheva, "A Theoretical Study on Six Classifier Fusion Strategies". IEEE transactions on pattern analysis and machine intelligence. Vol.24 NO.2, February 2002.

144. Z. Ma and A. Leijon, "Beta mixture models and the application to image classification". in Proceedings of IEEE International Conference on Image Processing, 2009.

145. Z. Ma and A. Leijon, "Human skin color detection in RGB space with Bayesian estimation of beta mixture models". Sound and Image Processing Lab KTH - Royal Institute of Technology, Aalborg, Denmark, August 23-27, 2010.

146. J. MacQueen, "Some methods for classification and analysis of multivariate observations". Dans Proceedings of the Fifth Berkeley Symposium on Mathematics, Statistics and Probability, Vol. 1, pages 281–296, 1967.

147. S. Marcel, O. Bernier et D. Collobert, "Approche EM pour la construction de

régions de teinte homogènes: application au suivi du visage et des mains d'une personne". Université de Poitiers, CORESA 2000, Octobre 2000.

148. J.A. Marcial-Basilio, G. Aguilar-Torres, G. Sánchez-Pérez, L.K. Toscano-Medina and H.M. Pérez-Meana, "Detection of Pornographic Digital Images". Internationnal journal of computers Issue 2, Volume 5, 2011.

149. B. Menser and M. Wien, "Segmentation and tracking of facial regions in color image sequences". In Proc. SPIE Visual Communications and Image Processing 2000, pp. 731–740, 2000.

150. K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors". In Proceedings of the Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, USA, June 2003.

151. S. Mika, G. Ratsch, J. Weston, B. Scholkopf, A.J. Smola and K.R. Mueller, "Constructing descriptive and discriminative non-linear features: Rayleigh coefficients in kernel feature spaces", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004.

152. S. Ming-Jung, V. Deepthi and V.K. Asari, "Neural network based skin color model for face detection". Proceedings of the 32nd Applied Imagery Pattern Recognition Workshop, Oct. 15- 17, IEEE Computer Society, USA., pp: 141-145, 2003.

153. L. Mostafa and S. Abdelazeem, "Face detection based on skin color using neural network", 2005. *GVIP 05 Conference, CICC, Cairo, Egypt*.

154. N. Mottin, "Localisation de visages dans des images acquises avec différents cadrages de caméras". Application à l'indexation, DEA ARAVIS (image et vision). Université Sophia Antipolis, Juin 2000.

155. T.A. Mysliwiec, "A freehand computer pointing interface". Technical report, University of Illinois at Chicago, 1994.

156. J.P. Nakache, J. Confais, Statistique Explicative Appliquée, Partie 2, "Modèle Logistique". pages 77-168, Technip, 2003.

157. Y.I. Ohta, T. Kanade, and T. Sakai. "Color information for region segmentation". Computer Graphics and Image Processing, 13: 222_241, 1980.

158. N. Oliver, A. Pentland, and F. Berard, "Lafter: Lips and face real-time tracker with facial expression recognition". In Proc of IEEE Conf. on Computer Vision and Pattern Recognition, pages 123_130, 1997.

159. E. Parzen, "On estimation of a probability density function and mode". Ann. Math. Stat., Vol.33, pp. 1065-1076, 1962.

160. K. Pearson, "On lines and planes of closest fit to systems of points in space". Philosophical Magazine, 2(6): 559–572, 1901.

161. P. Peer, J. Kovac, and F. Solina, "Human skin colour clustering for face detection". International Conference on Computer as a Tool, EUROCON 2003, Ljubljana, Slovenia, September 2003.

162. A. Pentland, B. Moghaddam and T. Starner, "View-based and modular eigenspaces for face recognition". IEEE-C-CVPR, Seattle, WA, USA, pp. 84-91, 1994.

163. M. Philip, M. Long, A. Rocco, A. Servedio and H. Ulrich Simon, "Discriminative Learning can Succeed where Generative Learning Fails". Information Processing Letters archive Volume 103 Issue 4, August, 2007.

164. S.L. Phung, D. Chai and A. Bouzerdoum, "A universal and robust human skin color model using neural networks". Proceedings of the International Joint Conference on Neural Networks, July 2001, IEEE Xplore Press, Washington, DC., USA.

165. S.L. Phung, A. Bouzerdoum, D. Chai, "A novel skin color model in ycbcr color space and its application to human face detection". In IEEE International Conference on Image Processing (ICIP'2002), vol. 1, pp. 289–292, 2002.

166. V. Planchon, "Traitement des valeurs aberrantes: concepts actuels et tendances générales". Biotechnol. Agron. Soc. Environ. 9 (1), 19–34, 7 Janvier 2005.

167. S.J. Press, S. Wilson, "Choosing between logistic regression and discriminant analysis". J.Amer.Stat.Assoc.73, 699-705, 1978.

168. F. Prêteux, P. Horain, H. Ouhaddi, and M. Preda, "Report on Core Experiment 3 on Hand Baps interpretation". ISO/IEC JTC1/WG1 MPEG97/M3332, March 1998, http://www-sim.int-evry/Publications.

169. F.K.H. Quek, T. Mysliwiec, and M. Zhao, "A freehand pointing interface". Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition, Zurich, Switzerland, pp. 372-377, June 1995.

170. R. Rakotomalala, "Pratique de la Régression Logistique: Régression Logistique Binaire et Polytomique". Version 2.0, Université Lumière Lyon 2, 27 Décembre 2009.

171. B.D. Ripley, "Pattern Recognition and Neural Networks", University Press, Cambridge, 1996.

172. A. Rosenfeld, "From image analysis to computer vision: An annotated bibliography". 1955-1979. Computer Vision and Image Understanding, 84: 298_324, 2001.

173. P.L. Rosin and T. Ellis, "Image difference threshold strategies and shadow detection". In 6th British Machine Vision Conference, pp. 347–356, 1995.

174. S. Roux and E. Petit, "Codeur H.263 amélioré par la visiophonie mobile". 7ème journée d'échange: Compression et représentation des signaux audiovisuels (CORESA), Dijon, Novembre 2001.

175. Y.D. Rubinstein and T. Hastie, "Discriminative vs. informative learning". In A. Press, editor, In Proc. of the Third International Conference on Knowledge and Data Mining, pages 49–53, 1997.

176. T.P. Ryan, "Some issues in logistic regression". Comm.Stat.-Theo. Meth. 29, 2019-2032, 2000.

177. E. Saber and A.M. Tekalp, "Frontal-view face detection and facial feature extraction using color, shape and symetry based cost functions". Pattern Recognition Letters, 19: 669_680, 1997.

178. G. Saporta, "Probabilités analyse des données et statistique". Technip, 1990.

179. D. Saxe and R. Foulds, "Toward robust skin identification in video images". 2nd International Face and Gesture Recognition Conference, Septembre 1996.

180. B. Schiele and A. Waiebel, "Gaze tracking based on face color". Proc. Int. Work. on automatic Face and Gesture Recognition, pp. 344-349, Zurich, 1995.

181. B. Scholkopf et A. J. Smola, "Learning with Kernels", MIT Press, 2002.

182. R. Schumeyer, and K. Barner, "A color-based classifier for region identification in video". In Visual Communications and Image Processing 1998, SPIE, vol. 3309, pp. 189–200, 1998.

183. K. Schwerdt and J.L. Crowley, "Robust Face Tracking using Color". In Proceeding of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, Mars 2000.

184. D.D. Sidibe, "Une technique de relaxation pour la mise en correspondance d'images". Thèse de doctorat en informatique à l'université de Montpellier2, 07 Décembre 2007.

185. L. Sigal, S. Sclaroff and V. Athitsos, "Estimation and prediction of evolving color distributions for skin segmentation under varying illumination". In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol. 2, pp. 152–159, 2000.

186. H. Simon, "Why should machines learn?". In R.S. Michalski, J.G. Carbonnel, and T.M. Mitchell, editors, Machine Learning: An Artificial Intelligence Approach. Morgan Kaufmann, 1983.

187. J.R. Smith and S.F. Chang, "Automated image retrieval using color and texture". Technical Report CU/CTR 408-95-14, Columbia University, July 1995.

188. N. Sokolovska, "Contributions à l'estimation de modèles probabilistes

discriminants: apprentissage semi-supervisé et sélection de caractéristiques". Thèse de doctorat en informatique de Télécom ParisTech, 25 Février 2010.

189. M. Soriano, B. Martinkauppi, S. Huovinen and M. Laaksonen, "Skin detection in video under changing illumination conditions". In Proc. Computer Vision and Pattern Recognition, vol. 1, pp. 839-842, 2000.

190. H. Steinhaus, "Sur la division des corps matériels en parties". Bulletin L'Académie Polonaise des Sciences, 4: 801–804, 1956.

191. H. Stoppiglia, "Méthodes statistiques de sélection de modèles neuronaux; applications financières et bancaires". Thèse de Doctorat de l'Université Pierre et Marie Curie - Paris VI, Décembre 1997.

192. Y. Sumi and Y. Ohta, "Detection of face orientation and facial components using distributed appearance modeling". Proc. Int. Work. on Automatic Face and Gesture Recognition, pp. 254-259, Zurich, 1995.

193. M.J. Swain and D.H. Ballard, "Color Indexing". International Journal of Computer Vision, 1998.

194. M. Talibi Alaoui, R. Touahni et A. Sbihi, "Classification des Images Couleurs par association des Transformations Morphologiques aux Cartes de Kohnen". pp.83-90, CARI 2004.

195. M. Tao, Z. Pen and G. Xu, "The Feature of Human Skin Color. Journal of Software". 12(7), 2001.

196. D.S. Tauban and W. Marcellin, "JPEG2000: Image Compression Fundamentals, Standards and Practice". Kluwer Academic Publishers, 2001.

197 M.R. Teague, "Image analysis via the general theory of moments". Journal of the Optical Society of America, 70(8), pp. 920–930, 1980.

198. M. Tenenhaus, "Statistique - Méthodes pour décrire, expliquer et prévoir, Chapitre 11, La régression logistique binaire", pages 387-460 ; Chapitre 12, "Régression logistique multinomiale: réponses polytomique et ordinale", pages 461-499, Dunod, 2007.

199. J.C. Terrillon, M. Shirazi, H. Fumakachi and S. Akamatsu, "Comparative performance of different skin chrominance models and chrominances spaces for the automatic detection of human faces in color images". In Proc. IEEE International Conference on Face and Gesture Recognition, Grenoble, pp. 54-61, March 2000.

200. G. Thibault, "Indices de formes et de textures: de la 2D vers la 3D Application au classement de noyaux de cellules". Thèse de doctorat en sciences, Aix Marseille Université, 18 Juin 2009.

201. A.N. Tikhonov et V. Y. Arsenin, "Solutions of ill-posed problems". John Wiley,

New York, 1977.

202. R. Tomassone, M. Donzart, J.J. Daudin, J.P. Masson, "Discrimination et classement". Paris, Masson, 172P, 1988.

203. K. Toyama. '"Look, Ma - No Hands !' hands-free cursor control with real-time 3d face tracking". In Proc. Workshop on Perceptual User Interfaces (PUI'98), 1998.

204. K. Toyama, J. Krumm, B. Brumitt and B. Meyers, Wallflower: Principles and practice of background maintenance. In Proceedings of the International Conference on Computer Vision (ICCV'99), pp. 255–261, 1999.

205. A. Trémeau et al., "Image couleur: de l'acquisition au traitement". Collection Sciences Sup, 480 pages, Dunod éditions, ISBN 2 10 006843 1, 2004.

206. Z. Tu, "Learning generative models via discriminative approaches". In IEEE Computer Vision and Pattern Recognition, 2007.

207. N. vandenbroucke, "Segmentation d'images couleur par classification de pixels dans des espaces d'attributs colorimétriques adaptés. Application à l'analyse d'images de football". Thèse de doctorat en Automatique et Informatique Industrielle à l'université de Lille1, 14 Décembre 2000.

208. V. Vapnik, "The Nature of Statistical Learning Theory". Springer, N.Y, 1995.

209. V. N. Vapnik, "Statistical Learning Theory". Wiley, 1st edition, 1998.

210. V.N. Vapnik, "The nature of statistical learning theory". Springer-Verlag, New York, NY, 1995.

211. V.N. Vapnik, "The Nature of Statistical Learning Theory". (Information Science and Statistics), Springer, 1999.

212. V. Vezhnevets, V. Sazonov, and A. Andreeva, "A survey on pixel-based skin color detection techniques". In Proc. 13th International Conference on the Computer Graphics and Vision, Moscow, Russia, pp. 85-92, September 2003.

213. R. Vijayanandh and G. Balakrishnan, "Skin Region Detection Using Hillclimbing Segmentation with Fuzzy C-Means Algorithm". Published in International Journal of Advanced Engineering & Application, Jan 2011 Issue.

214. J. B. Waite and W.J. Welsh, "An application of active contour models to head boundary location". Proc. British Machine Vision Conf., pp. 407-412, Oxford, 1990.

215. C. Wang and M. Brandstein, "Multi-source face tracking with audio and visual data". In IEEE MMSP, pp.169–174, 1999.

216. P. Wellner, "Interacting with Paper on the DigitalDesk". Rapport technique

EPC-93-110, EuroPARC, Xerox Center, 1993.

217. C. R. Wren, A. Azarbayejani, T. Darrell and A. Pentland, Pfinder: Real-time tracking of the human body. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), pp. 780–785, 1997.

218. H. Wu, T. Yokoyama, D. Pramadihanto and M. Yachida, "Face and facial feature extraction from color images". International Conference on Automatic Face and Gesture Recognition, pp. 345-350, October 1996.

219. A. Wu, M. Shah, and N. da Vitoria Lobo. "A virtual 3d blackboard: 3d finger tracking using a single camera". In Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), March 2000.

220. J. Yang and A. Waibel, "Tracking Human Faces in Real-Time". School of Computer Science, Technical report, Carnegie Mellon University, November, 1995.

221. J. Yang, W. Lu, and A. Waibel, "Skin color modeling and adaptation". In proceeding of ACCV'98, Vol.2, pp. 687-694, Hong Kong, 1998.

222. M.H. Yang, D.J. Kriegman and N. Ahuja, "Detecting faces in images: A survey". IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(1): 34_58, January 2002.

223. J. Yang, C. Liu and L. Zhang, "Color space normalization: Enhancing the discriminating power of color spaces for face recognition". Pattern Recognition. Volume 43, Issue 4. Pages 1454-1466, April 2010.

224. K. C. Yow et R. Cipolla, "Enhancing human face detection using motion and active contours". Proc. 3rd Asian Conf. on Comp. Vision, Vol. 1, pp. 515-522, Hong Kong, 1998.

225. A.A. Zaidan, H.A. Karim, N.N. Ahmad, G.M. Alam et B.B. Zaidan, "A new hybrid module for skin detector using fuzzy inference system structure and explicit rules". International Journal of the Physical Sciences Vol. 5(13), pp. 2084-2097, 18 October, 2010.

226. B.D. Zarit, B.J. Super and F. K.H. Quek, "Comparison of five color models in skin pixel classification". In Proceedings. International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, Corfu, Greece, pp.58-63, September 1999.

227. A. Zelinksy and J. Heinzmann, "Real time visual recognition of facial gestures for human computer interaction". 2nd Int. Conf. On Automatic Face and Gesture Recognition, pp. 351-356, Vermont, 1996.

228. W. Zhao, R. Chellappa, A. Ronsenfeld, and P. J. Phillips, "Face recognition: A literature survey". ACM Computing Surveys, pages 399_458, 2003.

229. H. Zheng, M. Daoudi and B. Jedynak, "Blocking Adult Images Based on Statistical Skin Detection". Electronic Letters on Computer Vision and Image Analysis 4(2): 1-14, 2004.

230. Q. Zhu, C.T. Wu, K.T. Cheng and Y.L. Wu, "An adaptive skin model and its applications to objectionable image filtering". Proceedings of the 12th Annual ACM International Conference on Multimedia. ACM Press, New York, USA., pp: 56-63. DOI: 10.1145/1027527.1027538, Oct. 10-16 2004.

231. J. Zhu and T. Hastie, "Kernel logistic regression and the import vector machine". Dans Proceedings of the 14th Conference on Advances in Neural Information Processing Systems (NIPS), pages 1081–1088. MIT Press, 2001.

232. J. Zhu and T. Hastie, "Kernel Logistic Regression and the Import Vector Machine". Journal of Computational and Graphical Statistics, Volume 14, Number 1, Pages 185–205, 2005.

233. D.A. Zighed et R. Rakotomala, "Extraction de connaissances à partir de données (ECD)". Techniques de l'Ingénieur. HA. 2002.