



Numéro d'ordre :4244

Université des Sciences et Technologies de Lille  
Ecole Doctorale Régionale Sciences pour l'Ingénieur Lille-Nord de France - 072  
Laboratoire d'Informatique Fondamentale de Lille  
Doctorat en Informatique  
Convention de co-tutelle avec l'Université Mohammed V Agdal Rabat

# Modèles probabilistes indexés par les arbres : Application à la détection de la peau dans les images couleur

par :

**Sanaa EL FKIH**

Thèse en co-tutelle dirigée par :

M. M. DAOUDI Professeur, TELECOM Lille 1, Lille France  
M. D. ABOUTAJDINE Professeur, FS Rabat, Maroc

Thèse soutenue le 20 Décembre 2008 devant le jury composé de :

M.	M. DAOUDI	Professeur, TELECOM Lille 1, Lille France	(Directeur)
M.	D. ABOUTAJDINE	Professeur, FS Rabat, Maroc	(Directeur)
M.	A. DEL BIMBO	Professeur, Université de Florence, Italie	(Rapporteur)
M.	R. OULAD HAJ THAMI	Professeur, ENSIAS, Rabat Maroc	(Rapporteur)
M.	D. HAMAD	Professeur, ULCO, Calais France	(Examinateur)
Mme.	S. MOULINE	Maître de Conférences, FS Rabat, Maroc	(Examinatrice)

Université des Sciences et Technologies de Lille  
Laboratoire d'Informatique Fondamentale de Lille  
UMR USTL/CNRS 8022 - Bat. M3  
Cité Scientifique - 59655 Villeneuve d'Ascq Cedex - France  
Tel: +33 (0)3 28 77 85 41  
Fax: +33 (0)3 28 77 85 37

Université Mohammed V Agdal  
Faculté des Sciences de Rabat  
4 Avenue Ibn Battouta  
B.P. 1014 RP, Rabat - Maroc  
Tel : +212 (0) 37 77 18 34  
Fax : +212 (0) 37 77 42 61





---

## AVANT PROPOS

Le travail de cette thèse a été réalisé au sein du Laboratoire d'Informatique Fondamentale de Lille (LIFL) de l'Université des Sciences et Technologies de Lille, France. Il a été effectué dans le cadre du programme de la co-tutelle, en collaboration avec le Laboratoire de Recherche en Informatique et Télécommunications (LRIT) de la Faculté des Sciences de Rabat Université Mohammed V Agdal, Maroc.

Je tiens tout d'abord à exprimer ma profonde gratitude à Monsieur Mohamed DAOUDI, professeur à TELECOM Lille 1, pour son aide précieuse, les efforts qu'il a prodigués pour l'accomplissement de ce travail, ainsi que pour la qualité de l'encadrement qu'il m'a assuré et sa rigueur scientifique.

J'exprime ma profonde reconnaissance à Monsieur Driss ABOUTAJDINE, professeur à la Faculté des Sciences de Rabat et responsable du LRIT, pour m'avoir encadré avec un intérêt constant et une grande compétence, pour sa disponibilité, son soutien, ses précieux conseils, et les encouragements qui m'ont permis de mener à bien ce travail.

Que mes directeurs de thèse soient assurés de toute mon estime et de mon plus profond respect.

Que Monsieur Denis HAMAD, professeur à l'Université du Littoral Côte d'Opale France, trouve ici l'expression de mes remerciements les plus sincères d'avoir accepté de présider cette thèse.

Je suis très honorée par la présence de Monsieur Alberto DEL BIMBO, professeur à l'Université de Florence Italie, et Monsieur Rachid OULAD HAJ THAMI, professeur à

l'École Nationale Supérieure d'Informatique et d'Analyse des Systèmes. Qu'ils trouvent ici mes plus vifs remerciements d'avoir accepté de juger ce travail et d'en être les rapporteurs. Mes sincères remerciements vont également à Monsieur Rachid BENSLIMANE, professeur à l'Université Sidi Mohamed Ben Abdellah Fès, Madame Salam MOULINE, professeur assistant à la Faculté des Sciences de Rabat, et Monsieur Abdennbi BELHADJ, chef du Service Veille Technologique Innovation & nouveaux services à Maroc Telecom, pour leur participation au jury de cette thèse.

J'exprime également mes remerciements aussi bien au Ministère des Affaires Etrangères Français de m'avoir accordé la bourse d'excellence Eiffel Doctorat, qu'à MAROC TELECOM qui a financé cette thèse.

J'adresse mes remerciements à Monsieur Bruno JEDYNAK, professeur à l'Université Johns-Hopkins, pour les discussions fructueuses que nous avons eues.

Mes remerciements vont aussi à tous mes collègues dans les laboratoires LRIT et LIFL. En particulier, un grand merci à Madame Fadoua ATAA-ALLAH et à Monsieur Benayad NSIRI, pour leurs conseils.

Finalement, je remercie chaleureusement mon mari, ma famille et ma belle famille pour leur soutien.

*à la mémoire de ma mère,  
à mon mari,  
à tous ceux qui m'ont supportés durant mes années des études*





---

## RÉSUMÉ

La détection de la peau constitue une phase primordiale de prétraitement dans plusieurs applications telles que la vidéo surveillance et le filtrage d'Internet. Toutefois, c'est une tâche difficile à accomplir étant donné la diversité des couleurs de la peau et la variété des conditions de prise de vues. Dans l'objectif de surmonter ces dernières contraintes, nos travaux de thèse consistent à définir un modèle robuste de la distribution de la peau capable de différencier les pixels de peau de ceux de non-peau dans des images variées.

Notre modélisation est fondée sur le concept des modèles graphiques probabilistes connus par leur intuitivité et efficacité pour la représentation d'une distribution jointe de probabilités sur un ensemble de variables aléatoires, plus particulièrement les arbres indexant des probabilités. En vue de définir le modèle de l'arbre idéal indexant la distribution de la peau, nous avons proposé trois approches différentes : le modèle d'arbre de dépendances à bi-probabilité peau et non peau, le modèle de mélange des arbres et celui de leur combinaison.

Le modèle d'arbre de dépendances à bi-probabilité peau et non peau proposé, exploite les propriétés d'interclasse et d'intra classe entre les deux classes peau et non peau ainsi que les interactions entre un pixel et ses voisins que nous traduisons par un arbre de dépendance optimal. L'arbre élaboré est un arbre idéal unique indexant conjointement les distributions de probabilités peau et non peau.

Le modèle de mélange des arbres est proposé pour remédier à la multiplicité des arbres de dépendances optimaux possibles sur un graphe. L'entité du mélange proposée concerne

aussi bien les structures des arbres considérés que les probabilités portées par ces dernières. Ainsi, l'arbre idéal indexant la probabilité peau est l'arbre résultant du mélange portant la probabilité du mélange.

Quant au modèle de combinaison des arbres élaboré, il constitue une approche alternative au mélange proposé visant l'exploitation des différentes informations emmagasinées dans les différents arbres de dépendances optimaux possibles. Un fondement théorique est présenté dans cette thèse pour déterminer la meilleure approche à adopter, le mélange des arbres ou la combinaison des arbres, et cela en fonction des arbres de dépendances optimaux considérés.

Les expérimentations réalisées sur la base Compaq montrent l'efficacité et la faisabilité de nos approches. En outre, des études comparatives entre nos modèles de peau et l'existant prouvent qu'en termes de qualité et de quantité des résultats obtenus, les modèles proposés permettent de discriminer les pixels de peau et ceux de non peau dans des images couleurs variées.

**Mots clés :** Classification, segmentation, détection de la peau, modèles graphiques probabilistes, arbre recouvrant optimal



---

## ABSTRACT

Skin detection or segmentation is considered as an important preliminary process in a number of existing systems ranging over face detection, filtering Internet images, and diverse human interaction areas. Nevertheless, there are two skin segmentation challenges : the pattern variability and the scene complexity. This thesis is devoted to define a new approach for modeling the skin probability distribution.

In the aim of dealing with the skin detection problem, we investigate the models of probability trees to approximate skin and non-skin probabilities. These models can represent a joint distribution in an intuitive and efficient way. Hence, we have proposed three main approaches to seek a perfect tree model estimating the skin probability distribution : (1) the model of dependency tree that approximates the skin and the non skin probability distributions together, (2) the mixture of trees' model, and (3) the combination of trees' model.

The first proposed model is based on the optimal spanning tree principle combined to an appropriate relevant criterion that we have defined. This contribution takes into account both the interclass and the intra class between skin and non skin classes, and the interactions between a given pixel and its neighbors.

The rationale behind proposing the second model is that in some cases the approximation of true class probability given by an optimal spanning tree (OST) is not unique and might be chosen randomly, while this model will take the advantages of the useful information represented on each OST. The mixture of trees' model consists in mixing the structures

of the OSTs and their probabilities with the aim of seeking a perfect spanning tree. This latter emphasizes the dependencies' degrees of data, and approximates effectively the true probability distribution.

Finally, the third model is defined to deal with a particular kind of multiple OSTs. This model is a parallel combination of different classifiers based on the OSTs.

A mathematical theory, proving and specifying the appropriate approach to be used (mixture of trees or combination of trees) depending on the considered OSTs' kind, is presented in this thesis. In addition to experimental results, on the Compaq database, showing the effectiveness and the high reliability of our three approaches.

**Key words :**Classification, segmentation, skin detection, graphical models, optimal spanning tree.



---

## TABLE DES MATIÈRES

Avant-Propos . . . . .	i
Table des matières . . . . .	xii
Table des figures . . . . .	xvi
Liste des tableaux . . . . .	.xvii
<b>Introduction générale . . . . .</b>	<b>1</b>
<b>1 Détection de la peau dans les images couleurs . . . . .</b>	<b>9</b>
1.1 Problématique de la détection de la peau . . . . .	10
1.2 Espaces de couleur pour la détection de la peau . . . . .	12
1.3 Modélisation de la distribution de la peau par la couleur . . . . .	16
1.3.1 Méthodes non paramétriques . . . . .	16
1.3.2 Méthodes paramétriques . . . . .	21
1.3.3 Seuils fixés . . . . .	24
1.4 Autres modèles pour la distribution de la peau . . . . .	24
1.5 Evaluations des modèles existants . . . . .	26
1.6 Discussion . . . . .	28
<b>2 Modèles probabilistes indexés par les arbres . . . . .</b>	<b>31</b>

2.1	Modèles graphiques probabilistes : Notions de bases . . . . .	32
2.2	Distributions d'arbres . . . . .	35
2.3	Arbres à structures prédéfinies et fixes . . . . .	38
2.3.1	Arbre 4-Star . . . . .	38
2.3.2	Arbre de Cayley . . . . .	39
2.4	Apprentissage de la structure d'un arbre . . . . .	41
2.4.1	Formulation du problème . . . . .	41
2.4.2	Résolution du maximum de vraisemblance . . . . .	42
2.4.3	Melange d'arbres . . . . .	44
2.5	Discussion . . . . .	46
<b>3</b>	<b>Vers de nouvelles distributions d'arbres indexant des probabilités . . . .</b>	<b>49</b>
3.1	Quelques notations . . . . .	50
3.2	Un nouveau modèle d'arbres pour indexer des probabilités . . . . .	51
3.2.1	Principe de base et motivation . . . . .	51
3.2.2	Modèle des arbres proposé . . . . .	53
3.3	Arbre de dépendances à bi-probabilité peau et non peau Simple . . . . .	55
3.3.1	Critère d'optimisation . . . . .	55
3.3.2	Modèle de l'arbre de dépendances à bi-probabilité peau et non peau Simple . . . . .	56
3.4	Pour un Meilleur Arbre Recouvrant . . . . .	58
3.4.1	Critère d'optimisation . . . . .	58
3.4.2	Modèle du Meilleur Arbre Recouvrant . . . . .	60
3.5	Vers un modèle d'arbres à bi-probabilité peau et non peau Discriminatoire . . . . .	63
3.5.1	Comparaison des deux modèles d'arbres à bi-probabilité peau et non peau proposés . . . . .	63
3.5.2	Modèle d'arbre de dépendances à bi-probabilité peau et non peau Discrimatoire . . . . .	65
3.5.3	Sélection de l'arbre de dépendances à bi-probabilité peau et non peau Discriminatoire . . . . .	67
3.6	Arbres de dépendances à bi-probabilité peau et non peau . . . . .	70
3.6.1	Procédures . . . . .	70
3.6.2	Inférences . . . . .	71

---

3.7	Discussion . . . . .	72
<b>4</b>	<b>Modèles probabilistes indexés par des mélanges ou des combinaisons des arbres . . . . .</b>	<b>75</b>
4.1	Ensemble des arbres recouvrants optimaux . . . . .	76
4.1.1	Multiplicité des arbres de recouvrement optimaux d'un graphe . . . . .	76
4.1.2	Extension des arbres de recouvrement optimaux . . . . .	78
4.1.3	Sélection des arbres optimaux d'un graphe . . . . .	79
4.2	En quête de l'arbre recouvrant "idéal" . . . . .	80
4.2.1	De la multiplicité des arbres recouvrants optimaux vers l'arbre recouvrant "idéal" . . . . .	80
4.2.2	Etude théorique du mélange des arbres de Chow Liu . . . . .	82
4.2.3	Etudes théoriques du mélange de trois types d'arbres à bi-probabilité peau et non peau . . . . .	85
4.3	Mélange des arbres . . . . .	89
4.3.1	Modèle de mélange . . . . .	90
4.3.2	Estimations de paramètres du modèle de mélange . . . . .	92
4.3.3	Quelques résultats expérimentaux . . . . .	94
4.4	Pour une combinaison des arbres recouvrants optimaux . . . . .	95
4.4.1	Méthodes de combinaison . . . . .	95
4.4.2	Approche pour la combinaison des arbres recouvrants optimaux . . . . .	97
4.5	Discussion . . . . .	100
<b>5</b>	<b>Détection de la peau par les nouveaux modèles d'arbres probabilistes : Experimentations et résultats . . . . .</b>	<b>103</b>
5.1	Protocole expérimental . . . . .	104
5.1.1	Corpus . . . . .	104
5.1.2	Espace de couleur . . . . .	106
5.1.3	Vecteur des observations . . . . .	106
5.1.4	Histogramme . . . . .	107
5.1.5	Classifieurs à évaluer . . . . .	108
5.2	Première étude comparative . . . . .	109
5.3	Deuxième étude comparative . . . . .	114

5.4 Troisième étude comparative . . . . .	123
5.5 Discussion . . . . .	126
<b>Conclusion générale . . . . .</b>	<b>129</b>
<b>Annexe . . . . .</b>	<b>134</b>
<b>A Le mélange des arbres de Chow Liu : Théorèmes . . . . .</b>	<b>135</b>
A.1 Preuve du théorème 5 chapitre 4 . . . . .	135
A.2 Preuve du théorème 6 chapitre 4 . . . . .	137
<b>Bibliographie . . . . .</b>	<b>141</b>



---

## TABLE DES FIGURES

1.1	Rôle d'un détecteur de la peau. . . . .	10
1.2	Quelques variétés de la couleur de la peau. . . . .	11
1.3	Quelques objets ayant une couleur semblable à celle de la peau. . . . .	11
1.4	Représentation d'une image couleur. . . . .	13
1.5	Perceptron multicouches. . . . .	19
1.6	Dispositions des neurones du réseau de Kohonen. . . . .	20
2.1	Exemple d'un graphe orienté. . . . .	33
2.2	Exemple d'un arbre non orienté. . . . .	36
2.3	Exemple d'un arbre orienté. . . . .	37
2.4	Deux arbres 4-Star de profondeurs égales à 1 (a) et à 2 (b). . . . .	39
2.5	Construction d'un arbre de Cayley de profondeur égale à 2. (a) graphe initial. (b) arbre de Cayley déduit. . . . .	40
3.1	Exemples de systèmes de voisinages d'ordre $r$ (avec $r = 0, 2, 4$ ). . . . .	50
3.2	Principe de construction des arbres indexant les probabilités peau et non peau. . . . .	52
3.3	Principe de construction de l'arbre de dépendance à bi-probabilité peau et non peau. . . . .	54

3.4	Quelques entrées et sorties de deux détecteurs de peau. Les images de la deuxième et troisième colonne correspondent respectivement à celles obtenues par le détecteur fondé sur l'arbre de Chow Liu et celui à base de l'arbre de dépendances à bi-probabilité peau et non peau Simple. . . . .	58
3.5	Des entrée (première colonne à gauche) et des sorties de deux détecteurs de peau fondés sur deux types d'arbres de dépendances à bi-probabilité. Les images aux niveaux de gris de la deuxième et de la troisième colonne correspondent respectivement à celles du détecteur fondé sur l'arbre Simple et celui basé sur le Meilleur Arbre Recouvrant. . . . .	63
3.6	Représentation graphique de deux ensembles $A$ et $B$ . . . . .	66
3.7	Des entrée (première colonne à gauche) et des sorties de deux détecteurs de peau fondés sur deux types d'arbres de dépendances à bi-probabilité. Les sorties sur la deuxième et troisième colonne correspondent respectivement à celles du détecteur fondé sur le Meilleur Arbre Recouvrant et celui basé sur l'arbre Discriminatoire. . . . .	69
4.1	Un graphe de dépendances (en haut) et ces arbres de recouvrements optimaux (en bas). . . . .	76
4.2	Procédé du mélange des $K$ arbres recouvrants optimaux. . . . .	90
4.3	Exemple de construction du graphe du mélange à partir de trois arbres de recouvrement ou de dépendances optimaux. . . . .	91
4.4	Quelques entrées et sorties des modèles de mélange des arbres pour la détection de la peau. La deuxième colonne représente les images aux niveaux de gris du modèle où un seul arbre est considéré alors que la troisième colonne donne celles du modèle où deux arbres sont mélangés. . . . .	94
4.5	Quelques entrées et sorties des modèles fondés sur la combinaison des arbres. Images sur la deuxième colonne correspondent au modèle du Meilleur Arbre Recouvrant alors que les images de la troisième colonne correspondent à celui de la combinaison de trois Meilleurs Arbres Recouvrants possibles. . . . .	99
5.1	Quelques images de la base Compaq. . . . .	105

5.2	Notations utilisées pour les composantes du vecteur des observations issu d'un bloc $3 \times 3$ . . . . .	107
5.3	Quelques entrées et sorties des différents modèles comparés. De gauche à droite, les sorties qui sont des images aux niveaux de gris correspondes aux modèles fondés sur le modèle de base, 2-BDTs, 1-BDT, diff-BDT et 1-BDT-diff-BDT respectivement. . . . .	110
5.4	Courbes ROC des classifieurs basés sur le modèle de base, 2-BDTs, 1-BDT, diff-BDT et 1-BDT-diff-BDT respectivement. . . . .	112
5.5	Les quatre arbres de dépendances optimaux indexant la classe peau, obtenus par l'apprentissage. . . . .	115
5.6	Les quatre arbres de dépendances optimaux indexant la classe non peau, obtenus par l'apprentissage. . . . .	116
5.7	Les arbres de mélanges indexant la distribution de peau ou celle de non peau, obtenus par les trois mélanges considérés ( $K = 2, 3, 4$ ). . . . .	119
5.8	Quelques entrées et sorties des classifieurs dont chacun est basé sur un couple d'arbres de peau et non peau. Première colonne représente les images couleurs (entrées). Ensuite les images aux niveaux de gris obtenues par les différents classifieurs sont données. Les sorties des classifieurs sont ordonnées de gauche à droite selon l'ordre croissant des numéros des arbres considérés $T^K$ ( $K = 1, 2, 3, 4$ ). . . . .	120
5.9	Quelques entrées et sorties des classifieurs basés sur les mélanges d'arbres. La première colonne représente les images couleurs correspondant aux entrées des classifieurs. Les sorties des classifieurs sont ordonnées de gauche à droite selon le nombres d'arbres mélangés $\Theta^K$ ( $K = 2, 3, 4$ ). . . . .	121
5.10	Courbes ROC des classifieurs étudiés fondés sur les modèles de mélanges d'arbres. . . . .	122
5.11	Quelques entrées et sorties des classifieurs de peau fondé sur les combinaisons d'arbres. La première colonne représente les images couleurs correspondant aux entrées des classifieurs. La deuxième et la troisième colonne correspondent respectivement aux sorties des classifieurs fondés sur Comb-1-ArbreROpt et Comb-3-ArbreROpts. . . . .	124

5.12 Courbes ROC des classifieurs étudiés fondés sur les combinaisons des Meilleurs Arbres Recouvrants. . . . .	125
---	-----



---

## LISTE DES TABLEAUX

1.1	Comparaison entre quelques méthodes de la classification de la peau. . . . .	27
5.1	Ensemble d'arbres considérés par chaque mélange étudié. . . . .	117
5.2	Erreurs de tolérance considérées par les différents mélanges. . . . .	117
5.3	Tableau récapitulatif des résultats obtenus par les trois études comparatives menées. . . . .	127
5.4	Comparaison entre les modèles de détection de la peau proposés dans cette thèse. . . . .	131





---

## INTRODUCTION GÉNÉRALE

### Contexte

Compte tenu des moyens techniques considérables dont nous disposons aujourd'hui, une révolution ample et expéditive est connue aussi bien dans le domaine de l'informatique que dans le domaine des télécommunications. En particulier, grâce à l'évolution des méthodes de la compression d'image, de son et de la vidéo ainsi que la vitesse remarquable des microprocesseurs, un échange massif des données numériques textuelles, sonores et visuelles (exemples : texte, image fixe et animée, vidéo et audio) n'a cessé de s'accroître régulièrement et rapidement. D'ailleurs les médias numériques ont envahi notre quotidien. Plus précisément, Internet est devenu le réseau public mondial le plus communément utilisé dans les administrations, les écoles, les entreprises et chez soi. En fait, la communauté des internautes tend à se diversifier, toutes les tranches d'âge sont présentes sur la toile et leur nombre s'accroît d'une manière exponentielle.

Par ailleurs, Internet n'est pas un simple média ordinaire constituant un support de diffusion des messages, mais il offre plutôt la possibilité d'une multitude de relations interactives entre les internautes, entre l'internaute et le monde réel ou encore entre l'internaute et lui-même. Face à l'importance de l'utilisation d'Internet, la facilité d'accès et la diversité des services disponibles sur cette toile, une grande prise de conscience de l'importance du traitement et de l'analyse des informations véhiculées s'est avérée impérativement indispensable.

En outre, le progrès technologique touche tous les secteurs d'activités (hôpitaux, commerces, administrations, ...) pour créer un monde où les échanges de services entre personne et établissement deviennent informatisés. Là aussi, il paraît d'ores et déjà nécessaire de traiter le flux de données numériques dans le but d'assurer une meilleure gestion de celles-ci.

Toutefois, même si de nos jours le traitement du contenu textuel ou sonore numérique connaît un développement considérable en termes d'indexation, de classification, de recherche ou autres analyses, le traitement du contenu visuel numérique demande encore des efforts à investir afin qu'il soit au mieux organisé. C'est dans ce contexte que s'inscrit la présente thèse. Il s'agit ici de proposer et d'étudier des solutions pour le traitement du contenu numérique visuel plus particulièrement le traitement des images.

## Problématique et objectifs du mémoire

L'image est depuis toujours, considérée comme le support le plus riche et le plus expansif qui permet de remplacer toute expression. En effet, l'image reste le moyen de communication le plus efficace, le plus attirant et le plus approprié pour transmettre des messages entre différentes personnes quelque soit leur ethnie. Elle est dite le langage universel qui traduit les informations sans se restreindre à une communauté de personnes particulière. Ce sont principalement les raisons pour lesquelles le contenu visuel s'impose et s'utilise d'une manière exceptionnelle en particulier dans le web. La gestion du contenu visuel est indispensable pour une meilleure maîtrise des informations que l'utilisateur reçoit. La gestion la mieux adéquate comporte les différents traitements des images, à savoir l'indexation, la classification, la recherche, et globalement tout ce qui concerne l'analyse sémantique d'une image. Ce traitement peut être différencié en un traitement direct du contenu visuel de l'image et un traitement par analyse de son annotation. L'annotation d'une image contient des mots clés que l'utilisateur spécifie afin de décrire cette image. Toutefois, chaque utilisateur peut décrire le contenu d'une image à sa propre manière. Cela laisse le champ libre à tout utilisateur malintentionné d'introduire des spécifications de l'image qui ne sont pas réelles, et cela dans le but de corrompre toute analyse de l'image. D'ailleurs, c'est le cas de plusieurs internautes qui désirent faire circuler, sur la toile, des images illicites (images pour adultes, images exprimant du racisme, ...). L'analyse par

contenu visuel consiste à analyser chaque image selon le contenu réel qu'elle représente, loin de toute falsification.

Le traitement d'une image par analyse de son contenu visuel reste le moyen le plus efficace. Dans ce cadre, il existe plusieurs manières d'analyser le contenu visuel telle que la segmentation, l'extraction, la recherche, l'indexation, la classification, etc. Dans cette perspective, nous nous limitons à la classification des images. Les techniques de classifications se divisent en deux types : la classification supervisée et la classification non supervisée. Il y a aussi quelques auteurs qui parlent de classification semi-supervisée qui n'est autre qu'une combinaison entre les deux premiers types de classifications cités. En outre, dans une image, la classification peut bien concerner l'image entière ou bien des zones de celle-ci (ensemble de pixels). Nous nous intéressons à la dernière catégorie de la classification dans un cadre supervisé. Plus particulièrement, dans le cas d'une classification binaire, nous désirons classer les pixels d'une image en pixels peau et pixels non peau. L'objectif de cette classification réside dans la résolution de la problématique de la détection de la peau dans les images couleurs. En fait, comme sera présenté dans le chapitre 1, la détection de la peau dans une image est une tâche difficile à accomplir, toutefois il s'agit d'un prétraitement primordial dans plusieurs applications telles que le suivi des gestes de la main [4], la vidéo surveillance [6], le filtrage d'Internet [8] et la segmentation de la vidéo [10]. La sélection des zones peau et des zones non peau dans une image, permet de restreindre l'analyse de l'image entière à l'analyse des zones peau candidates de l'image qui sont susceptibles de contenir une ou plusieurs parties en quête du corps humain telles que les mains, le visage, les pieds et la silhouette de l'être humain. Une telle analyse constitue une étape fondamentale qui assure de minimiser la complexité des traitements s'articulant autour de l'étude d'une partie du corps humain. En outre, la problématique de la détection de la peau est d'une grande importance dans le domaine de la recherche scientifique et la littérature en est témoin. Cette dernière étale plusieurs techniques s'appuyant sur différentes caractéristiques pour résoudre la problématique en question, or l'information couleur reste le moyen le plus couramment utilisé à cette fin. D'ailleurs, différents espaces de couleurs et maintes méthodes paramétriques et non paramétriques sont utilisés pour la modélisation de la distribution de la couleur de la peau afin de séparer les pixels peau de ceux non peau dans les images couleurs (voir un panorama de ces méthodes exposé dans le chapitre 1). En dépit de cette diversité dans la

modélisation de la distribution de la peau, un modèle robuste pour discriminer les pixels de la peau dans les images couleurs est toujours sollicité. Cela est principalement l'objectif du présent mémoire. Plus particulièrement, nous optons pour un classifieur probabiliste qui marque le degré d'appartenance de chaque pixel d'une image, à la classe peau. Un tel classifieur permettra à son utilisateur de spécifier le seuil de discrimination des classes peau et non peau. Dans nos travaux nous adoptons les modèles graphiques probabilistes pour atteindre notre objectif.

Les modèles graphiques probabilistes sont définis comme étant le mariage entre la théorie des probabilités et la théorie des graphes. Ils permettent de prendre en compte l'aspect incertain présent dans les applications réelles et d'offrir un outil attractif pour la modélisation des dépendances entre les données. Cette utilisation conjointe des probabilités et des graphes nous offre une famille de modèles de connaissance très riche. Dans le présent mémoire, nous proposons d'introduire une nouvelle classe de modèles probabilistes pour l'analyse de données graphiques (images). Dans ce contexte, le choix est la plupart du temps entre des modèles réalistes mais d'utilisation extrêmement difficile comme les champs de Markov ou des modèles ultra-simplifiés qui supposent par exemple l'indépendance des pixels ou encore une dépendance Markovienne suivant un ordre arbitraire sur les pixels, tant il est difficile de concilier une structure de dépendance riche et une efficacité pratique.

En tenant compte du fait que dans une image, un pixel appartenant à la classe peau se trouve toujours dans une zone homogène de pixels peau, il est nécessaire de considérer le voisinage d'un pixel lors de sa classification en pixel peau ou pixel non peau. Ainsi, afin d'incorporer l'information de dépendance entre un pixel et son voisinage, nous utilisons les arbres probabilistes qui sont marqués par leur simplicité et efficacité dans la mesure où ils ne contiennent pas de cycles. Plus exactement, nous utilisons des arbres dits arbres de dépendances qui mettent en valeur les dépendances entre les données qu'ils représentent. Par le même biais, ces arbres permettent d'indexer des distributions de probabilités (dans notre cas les probabilités peau et non peau). Si dans le cadre de résolution de la problématique de la détection de la peau dans les images, des auteurs comme Zheng [15] ou Sebe [20] ont déjà eu recours dans leurs travaux à ce genre d'arbres présenté, l'originalité de nos travaux réside dans la proposition de nouveaux modèles pour la séparation de pixels peau et non peau, qui sont fondés sur le mélange ou la combinaison

des arbres de dépendances. Aussi, nous définissons des arbres de dépendances originaux, basés sur de nouveaux critères d'optimisation que nous jugeons être plus adéquats pour une modélisation au mieux des distributions de probabilités des classes peau et non peau. Par ailleurs, les expérimentations réalisées sur la plus large base d'images de peau Compaq [52] témoigneront de la validité et l'apport de nos propositions, en particulier vis-à-vis du modèle de base donné dans [86].

## Contributions et contenu du mémoire

La suite du présent mémoire est organisé comme ci-dessous :

- Le premier chapitre a pour fin d'élucider la problématique de la détection de la peau dans les images couleurs et de présenter quelques facteurs rendant la résolution de celle-ci aussi ardente. Il existe un champ fertile de travaux de recherches dédiés à la détection de la peau. Ce chapitre vise aussi à exposer un état de l'art sur la détection de la peau. Un panorama de méthodes paramétriques et non paramétriques utilisant différentes informations caractérisant les zones de peau dans une image, est exhibé. En outre, une évaluation des performances de quelques techniques de détection de la peau, est présentée.
- Le deuxième chapitre explicite le concept de modèles de probabilités indexés par les arbres ou ce qu'on appelle aussi les distributions d'arbres. Pour cela, des notions de base de la théorie des graphes sont données afin de se familiariser avec les terminologies qui seront utilisées dans ce mémoire. Les arbres indexant des probabilités peuvent être divisés en deux catégories, à savoir des arbres dont la structure est définie à priori et les arbres dont la structure est déterminée à partir des données du domaine étudié. Ces deux catégories sont décrites et illustrées par des exemples d'arbres communément utilisés.
- Le troisième et le quatrième chapitre sont consacrés à la description de nos propositions et nos visions pour l'élaboration de nouveaux modèles basés sur les arbres de probabilités visant la résolution de la problématique de la détection de la peau dans les images couleurs.

Le chapitre 3 exhibe nos propositions pour une génération d'arbres dite arbre de dépendances à bi-probabilité indexant conjointement et la probabilité peau et la proba-

bilité non peau. A travers trois différentes propositions des critères optimaux possibles caractérisant les distributions peau et non peau, trois modèles d'arbres originaux sont donnés pour la résolution de la problématique de la détection de la peau. Cette diversité de propositions est justifiée par notre but d'étudier les différentes possibilités pour aboutir à un modèle d'arbres modélisant au mieux la probabilité de peau et celle de non peau réelles. Ces différents arbres proposés sont fondés sur le principe d'arbre de recouvrement optimal défini sur un graphe.

Quant au chapitre 4, celui-ci exploite deux nouvelles pistes pour améliorer les modèles basés sur un arbre indexant des distributions de probabilités. En fait, sur un graphe donné il est possible que plusieurs arbres recouvrant optimaux coexistent. Cette coexistence impose la définition de l'arbre de recouvrement optimal "idéal" parmi tous ceux possibles. Au vu d'exploiter tous les apports des différents arbres sélectionnés, nous proposons dans ce chapitre le mélange de ces différents apports. Le mélange proposé concerne conjointement le mélange des différentes probabilités co-estimées par les arbres et le le mélange des structures de ces derniers. Des études théoriques sont présentées afin de prouver les améliorations apportées par les mélanges de différents genre d'arbres de probabilités. Dans des cas particuliers où la théorie ne permet pas de décider des apports des modèles de mélanges d'arbres vis-à-vis des modèles fondés sur un arbre du même type, une méthode de combinaison parallèle est donnée comme alternative du mélange proposé.

- Le cinquième chapitre aborde l'évaluation expérimentale des différents modèles de peau proposés vis-à-vis de modèles de bases connus. Dans cette perspective, trois études comparatives sont menées. La première étude évalue les modèles de peau fondées sur un arbre, proposés dans le troisième chapitre. Ces modèles sont comparés par rapport au modèle de base, donné dans [86], qui a offert les meilleures performances sur la base Compaq. Ils sont aussi comparés par rapport au modèle d'arbres de base défini dans [92]. La deuxième étude comparative est consacrée aux mélanges proposés des arbres alors que la troisième étude vise les combinaisons des arbres. A noter qu'un seul protocole expérimental défini au début de ce chapitre, est adopté dans les différentes méthodes et études comparatives.
- Pour conclure ce mémoire, une conclusion générale est donnée à la suite des cinq chapitres précédemment présentés. Celle-ci constitue un récapitulatif de nos travaux menés

dans le cadre de l'élaboration d'un modèle robuste de peau et présente des perspectives sur lesquelles s'ouvrent ceux-ci.

## Liste de publications

### \* Chapitre de livre

- S. El Fkihi, M. Daoudi, and D. Aboutajdine, Probability Approximation Using Best-Tree Distribution for Skin Detection, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Volume 4179/2006, Pages : 767-775, 2006.

### \* Journaux internationaux

- S. El Fkihi, M. Daoudi, and D. Aboutajdine, The Mixture of K-Optimal-Spanning-Trees Based Probability Approximation : Application to Skin Detection, Journal of Image Vision and Computing, February 2008.
- S. El Fkihi, M. Daoudi, and D. Aboutajdine, Discriminative Dependency Tree Based Skin and Non-Skin Probability Approximations, submitted to Pattern Recognition, November 2007 (second revision).

### \* Conférences internationales (avec comité de lecture et actes)

- S. El Fkihi, M. Daoudi and D. Aboutajdine. Optimal Spanning Trees Mixture based probability approximation for Skin Detection. In Proceedings of the second International Conference on Computer Vision Theory and Applications. Barcelona, Spain, March 8th - 11th, 2007.
- S. El Fkihi, M. Daoudi, and D. Aboutajdine, Probability Approximation Using Best-Tree Distribution for Skin Detection, Advanced Concepts for Intelligent Vision Systems, University of Antwerp, Antwerp, Belgium, September 18th - 21th, 2006.
- S. El Fkihi, M. Daoudi, and D. Aboutajdine, A Tree Distribution for Skin Detection. In Proceedings of the second International Symposium on Communications, Control and Signal Processing, Marrakech, Morocco, March 13th - 15th, 2006.
- S. El Fkihi, M. Daoudi and D. Aboutajdine, Un Algorithme basé sur les Meilleurs

Arbres de Dépendances pour la Détection de la Peau dans les Images Couleurs. In Proceedings of the 12èmes journées d'étude et d'échange COMpression et REprésentation des Signaux Audiovisuels, Montpellier, France, November 8th - 9th, 2007.

\* Conférences nationales (sans comité de lecture)

- S. El Fkihi, A. Jilbab, D. Aboutajdine, et M. Daoudi, Filtrage de sites à contenu illicite sur Internet. Journées doctorales LRIT (JDLRIT 08), Faculté de Sciences Agdal Université Mohamed V Rabat. 25-26 Avril 2008.

\* Rapports internes

- S. EL FKIHI et A. JILBAB. "Reconnaissance des Sites Illicites Basée sur la Détection des vidéos d'adultes". Rapport numéro 4, nombre de pages : 61. Projet "FILTRAGE DE SITES A CONTENU ILLICITE SUR INTERNET". 20 Juin 2008.
- S. EL FKIHI et A. JILBAB. "Reconnaissance des Sites Illicites Basée sur la Détection des images d'adultes". Rapport numéro 3, nombre de pages : 66. Projet "FILTRAGE DE SITES A CONTENU ILLICITE SUR INTERNET". 18 Juillet 2008.
- A. JILBAB et S. EL FKIHI. "Reconnaissance des Sites Illicites Basé sur la Détection des Logos". Rapport numéro 2, nombre de pages : 39. Projet "FILTRAGE DE SITES A CONTENU ILLICITE SUR INTERNET". 22 Juillet 2007.
- A. JILBAB et S. EL FKIHI."Etat de l'Art des Systèmes de Filtrage". Rapport numéro 1, nombre de pages : 67. Projet "FILTRAGE DE SITES A CONTENU ILLICITE SUR INTERNET". 27 Mars 2006.

---

## DÉTECTION DE LA PEAU DANS LES IMAGES COULEURS

La détection de la peau dans une image consiste à séparer les pixels de peau des pixels non peau. Elle constitue une étape de prétraitement fondamentale dans plusieurs systèmes opérant dans le cadre des applications dédiées aux interactions du corps humain. Cependant, c'est une tâche difficile à réaliser. Pour cela, plusieurs techniques ont été utilisées pour différencier entre les pixels peau et non peau dans une image.

Le présent chapitre a pour fin de présenter un état de l'art sur la détection de la peau dans les images couleurs. Dans une première section, la problématique de la détection de la peau dans les images sera présentée. Ensuite, dans les sections qui succèdent un panorama des travaux de recherches menés dans ce cadre, sera exposé.

Dans le domaine de la détection de la peau, les travaux de recherches peuvent se distinguer par les descripteurs sur lesquels ils sont basés. Or, l'information couleur est souvent utilisée comme la caractéristique de discrimination des deux classes peau et non peau. Pour ce, les méthodes de la détection de peau sont séparées en deux ensembles : un ensemble des méthodes exploitant la propriété couleur et un deuxième comportant des méthodes utilisant d'autres propriétés que la couleur. Nous présentons dans ce chapitre des échantillons de chaque ensemble. En outre, dans une approche basée couleur, deux axes de recherches sont abordés : le choix de l'espace de couleur et la méthode de modélisation de la couleur. Ainsi, un aperçu sur les différents espaces colorimétriques utilisés dans la détection de la peau, est présenté dans ce chapitre.

## 1.1 Problématique de la détection de la peau

La détection ou la segmentation de la peau humaine dans une image couleur consiste à détecter les pixels de la peau et ceux de non-peau dans cette image. Ainsi, Ayant une image couleur comme entrée, le rôle d'un détecteur de la peau se résume à fournir une image binaire discriminant les pixels de la peau des autres pixels. L'image en sortie du détecteur possède les mêmes dimensions que celles de l'image en entrée, et elle est appelée masque. La Figure 1.1 illustre le fonctionnement global d'un détecteur de la peau.

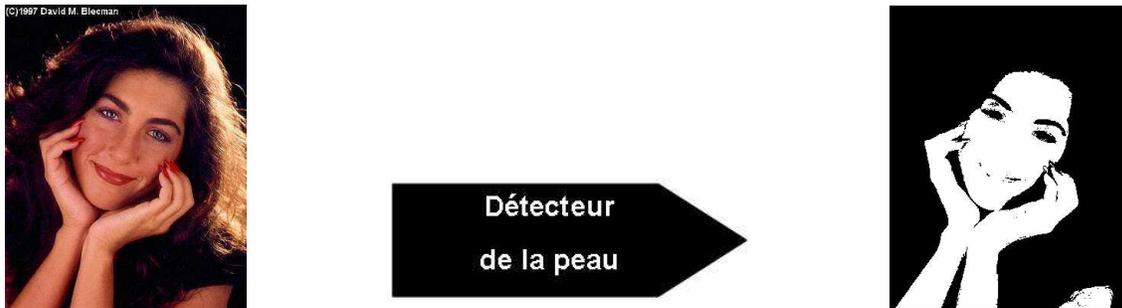


FIGURE 1.1 – Rôle d'un détecteur de la peau.

La détection de la peau dans une image permet de délimiter les zones contenant de la peau humaine. A partir de ce principe, elle est considérée comme un traitement préliminaire dans plusieurs problèmes de reconnaissances des diverses parties du corps humain tel que les mains [1], le visage [2] et toutes les parties nues du corps entier [3].

Un prétraitement basé sur la peau constitue une phase primordiale dans plusieurs systèmes appartenant à divers secteurs d'activités. En effet, il leur permet de focaliser leurs traitements sur des zones d'intérêts candidates potentielles de l'image. Par la suite, une garantie en performances en termes de temps de traitement du système ainsi que des fausses alarmes qui pourraient surgir par le traitement des pixels non-peau, est assurée.

Nous citons à titre d'exemple les systèmes de suivi des gestes de la main [4][5], de la vidéo surveillance [6][7], de filtrage d'Internet [8] [9] et de la segmentation de la vidéo [10][11].

Bien que la détection de la peau dans une image soit considérée comme une tâche facile à réaliser pour l'oeil humain, il est loin d'être le cas pour un détecteur automatique. En effet, la peau est une entité dynamique qui change sous l'influence de plusieurs facteurs tels que les variations de :

- la couleur de la peau (peau blanche, peau noire, peau jaune, ...) en fonction des différentes ethnies qui existent et qui sont déterminées par le mode de vie de différentes populations ;
- le niveau de la luminance de l'environnement où l'image est prise en vue ( par exemple : image intérieure ou extérieure ).
- paramètres du dispositif d'acquisition de l'image ( bruit, contraste, ... ) ;
- ...

Par ailleurs, les objets qui ont une couleur semblable à celle de la peau (bois, sable, charbon, ...) et qui se traduisent par un arrière plan complexe, influence aussi sur la définition de la couleur de la peau humaine.

Ces différents facteurs créent une classe de la peau avec beaucoup de variations interclasses comme l'illustre les Figures 1.2 et 1.3.



FIGURE 1.2 – Quelques variétés de la couleur de la peau.

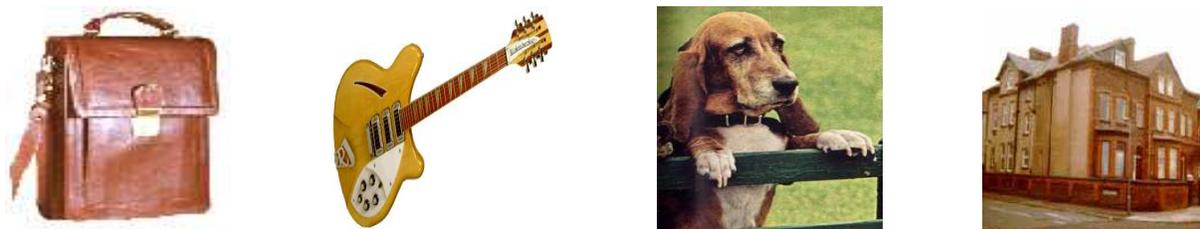


FIGURE 1.3 – Quelques objets ayant une couleur semblable à celle de la peau.

Du haut en bas, la première ligne de la Figure 1.2 montre que la couleur de la peau change d'une personne à une autre. Les images de la deuxième ligne présentent la peau dans différentes conditions de prise de vue. Quant aux images représentées à la Figure 1.3, elles donnent des exemples d'objets ayant des couleurs confondues avec celles de la peau. Les méthodes utilisées dans la détection de la peau sont regroupées en deux ensembles : (1) des méthodes basées sur l'information de la couleur et (2) des méthodes considérant autres propriétés que la couleur. Dans la suite du chapitre, nous présentons des exemples de ces deux ensembles. Toutefois, il est à noter que les approches utilisant l'information couleur, couvrent deux axes de recherches importants. Il s'agit du choix de l'espace de couleur le plus approprié et le deuxième axe concerne le modèle adéquat pour la modélisation de la distribution de la peau. Dans un premier temps, nous présentons les différents espaces de couleurs utilisés dans la détection de la peau.

## 1.2 Espaces de couleur pour la détection de la peau

La couleur est la primitive pertinente la plus répandue pour la séparation des pixels de la peau des pixels non-peau dans une image. La raison principale derrière le choix de ce descripteur peut être expliquée par le fait que, dans un espace de couleur, la gamme des couleurs de la peau est située dans une bande étroite comme le montre les travaux [12], [14] et [13]. De plus, la couleur est une information incontestable à utiliser grâce à sa robustesse vis à vis du problème de l'occultation, les transformations géométriques ou encore les changements d'échelles.

La colorimétrie, l'infographie et les normes de transmission de signal visuel ont donné naissance à plusieurs espaces de couleur. Une grande variété de ceux-ci a été utilisée afin de représenter la couleur de la peau. Cette multiplicité est justifiée par la délicatesse du choix de l'optimal repère pour une modélisation de la peau. Nous rapportons ici les espaces de couleur les plus populaires utilisés dans le cadre de la détection de la peau.

### Les espaces RVB et rvb

Le codage RVB (rouge, vert et bleu) mis au point en 1931 par la Commission Internationale de l'Eclairage (CIE), consiste à représenter l'espace des couleurs à partir de trois rayonnements monochromatiques de couleurs :

- rouge (de longueur d'onde égale à 700,0 nm) ;

- vert (de longueur d’onde égale à 546,1 nm) ;
- bleu (de longueur d’onde égale à 435,8 nm).

Cet espace de couleur correspond à la façon dont les couleurs sont généralement codées informatiquement, ou plus exactement à la manière dont les tubes cathodiques des écrans d’ordinateurs représentent les couleurs ; comme il est illustré par la Figure 1.4. Ainsi, dans le cadre général où la détection de la peau est une phase préliminaire d’une application fonctionnant en temps réel, l’espace RVB est souvent utilisé à cause de sa simplicité du moment que les pixels des images sont par défaut représentés dans cet espace. A titre d’exemple, les travaux [15], [16] et [17] ont utilisé l’espace de couleur RVB dans la modélisation de la peau.

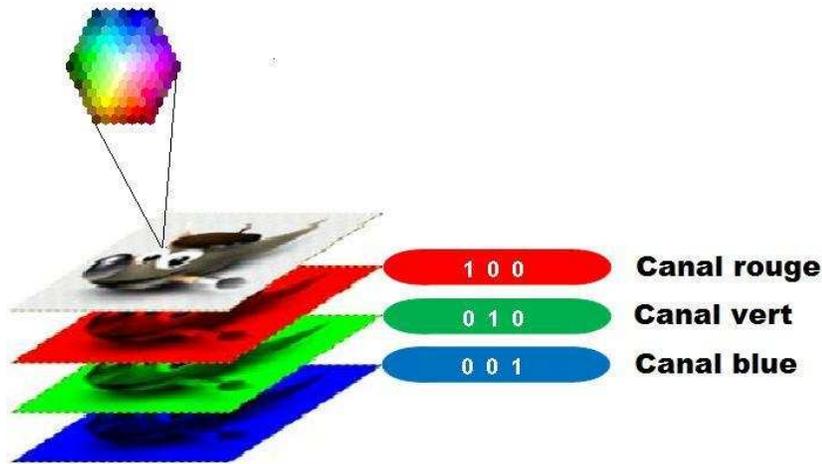


FIGURE 1.4 – Représentation d’une image couleur.

Une couleur est représenté par ces trois composantes (ou encore canaux) dans l’espace RVB. Or le triplet  $(R, V, B)$  ne représente pas que la couleur mais il encapsule en plus la luminosité. Cette dernière peut être supprimée par une normalisation des trois canaux comme suit :

$$r = \frac{R}{R + V + B} \quad v = \frac{V}{R + V + B} \quad b = \frac{B}{R + V + B}$$

d’où la naissance de l’espace de couleur  $rvb$  dit aussi RVB normalisé.

Comme il peut être constaté, la composante  $b$  peut être omise grâce à la propriété  $r + v + b = 1$ . Il vient donc qu’une couleur peut être représentée dans un espace à deux dimensions par le couple  $(r, v)$ . La réduction de la dimension et la simplicité de passer de l’espace RVB à

l'espace rvb, ont attiré l'attention de plusieurs chercheurs dans le cadre de la modélisation de la couleur de la peau, tel est le cas de [18], [19] et [19]. Toutefois, étant donnée que les valeurs des canaux  $r$  et  $v$  sont faibles (comprise entre 0 et 1) celles-ci sont plus sensibles au bruit.

### L'espace HSV

Le système (Hue-Saturation-Value) est défini par un cylindre qui représente la teinte, la saturation et la valeur d'une couleur. Cette dernière traduit l'intensité du flux lumineux perçu alors que la saturation explicite l'éloignement vis-à-vis de la lumière blanche c'est ce qu'on appelle la pureté relative de la couleur. Quant à la teinte ou hue, c'est un attribut associée à la longueur d'onde dominante dans la distribution poly chromatique perçue. D'autres variétés de cet espace sont :

- HSB : Hue, Saturation, Brightness soit Teinte, Saturation, Brillance en français. La brillance décrit la perception de la lumière émise par une surface.
- HSL : Hue, Saturation, Luminance soit Teinte, Saturation, Luminance en français.
- HSI : Hue, Saturation, Intensity soit Teinte, Saturation, Intensité en français.
- HCI : Hue, Chrominance, Intensity soit Teinte, Chrominance, Intensité.

L'espace HSV a été utilisé dans plusieurs travaux de recherches faisant appel à la détection de la peau tel que [21], [22], [23] et [24], grâce à la notion de séparation entre chrominance et luminance qu'il assure.

Le passage de l'espace RVB à l'espace HSV est non linéaire. Il est donné comme suit [25] :

$$Max = \max(R, V, B) \quad Min = \min(R, V, B)$$

$$Si \quad Max = Min \quad alors \quad H = 0$$

*Sinon*

$$Si \quad R = Max \quad alors \quad H = \left(0 + \frac{V - B}{Max - Min}\right)$$

$$Si \quad V = Max \quad alors \quad H = \left(2 + \frac{B - R}{Max - Min}\right)$$

$$Si \quad B = Max \quad alors \quad H = \left(4 + \frac{R - V}{Max - Min}\right)$$

$$Si \quad Max = 0 \quad alors \quad S = 0$$

$$Sinon \quad S = \frac{Max - Min}{Max}$$

$$V = Max$$

Le problème qui s'impose dans l'espace de couleur HSV est la singularité des valeurs S et H.

### Les espaces YUV et YCrCb

L'espace YUV, auparavant baptisée YCrCb, est utilisé par les anciens standards MPEG (MPEG 1, 2 et 4).

Plusieurs articles considèrent que c'est l'espace le plus adéquat pour la représentation de la couleur de la peau du fait qu'il sépare la luminance et la chrominance [26][27][28] [29] et [30]. En fait, le paramètre Y représente la luminance (c'est-à-dire l'information en noir et blanc), tandis que U et V permettent de représenter la chrominance, c'est-à-dire l'information sur la couleur. Ce modèle a été mis au point afin de permettre de transmettre des informations colorées aux téléviseurs couleurs, tout en s'assurant que les téléviseurs noir et blanc existant continuent de fonctionner.

La transformation de l'espace RGB en YCrCb est réalisée d'une manière linéaire via la matrice de transition suivante :

$$\begin{pmatrix} Y \\ Cr \\ Cb \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.500 & -0.419 & -0.081 \\ -0.169 & -0.331 & 0.500 \end{pmatrix} \begin{pmatrix} R \\ V \\ B \end{pmatrix}$$

### Autres espaces de couleur

Il existe d'autres espaces de couleur qui ont été utilisés dans la représentation de la couleur de la peau tel que YIQ [32] dans [33], Lab[34][35] [36] et CIE-Luv[37].

L'espace de couleur le plus simple à utiliser est l'espace RBG. Les autres espaces de couleur sont déterminés à partir des transformations linéaires ou non linéaires de l'espace RGB, et qui sont quelques fois coûteuses en termes de temps de transition.

Comme nous avons pu le constater dans les propos ci-dessus, certaines solutions essaient d'obtenir de meilleurs résultats dans la représentation de la peau, en exploitant la composante de la chrominance indépendamment de celle de la luminance. Toutefois, il est à noter que le changement d'espace de couleur induit des pertes d'informations.

Bien que la modélisation de la couleur de la peau dépend de l'espace de couleur utilisé [38], il a été rapporté dans [39] et [40] que le choix de l'espace de couleur devient secondaire lors de l'utilisation des méthodes non paramétriques ( que nous allons voir par la suite)

et d'une large base d'apprentissage.

## 1.3 Modélisation de la distribution de la peau par la couleur

Le principe de la classification est d'identifier les classes auxquelles appartiennent des objets à partir de traits descriptifs de ces derniers. Or, la classification peut être réalisée par deux manières différentes : supervisée et non-supervisée. Dans le cas où les classes sont connues et on dispose des échantillons de chaque classe dit encore vérité de terrain, il s'agit d'une classification supervisée. La classification non-supervisée est la contre-opposée de la précédente. Dans nos travaux nous adaptons une démarche supervisée pour la classification des pixels d'une image en classe peau ou non-peau. La vérité de terrain dont nous disposons est un ensemble des masques des images.

Un champ fertile de méthodes utilisées dans la détection de la peau et basées sur l'information de la couleur, est exposé dans la littérature. Des études et des synthèses sur les techniques de modélisations de la peau sont présentées dans [41], [42], [43], [44], [45]. Ces différentes méthodes ont pour objectif de déterminer la distribution de la classe peau et celle de la classe non peau. Elles sont divisées en trois sous ensembles : (1) l'ensemble des méthodes paramétriques, (2) l'ensemble des méthodes non paramétriques et (3) l'ensemble des méthodes basées sur des seuils fixés de définition de la peau. Nous présentons par la suite, des échantillons de chacun de ces ensembles.

### 1.3.1 Méthodes non paramétriques

Les méthodes non paramétriques déterminent les distributions des classes en se basant que sur les observations issues des différents échantillons des classes sans introduire aucune hypothèse sur les formes des distributions.

#### 1.3.1.1 Table normale de correspondance

Plusieurs algorithmes de détection de la peau utilisent des histogrammes pour segmenter les pixels de la peau. Ces histogrammes représentent les couleurs de l'image de traitement et ils sont construits à partir d'un ensemble des images d'apprentissage contenant les

échantillons de la peau sous ces différentes couleurs (blanche, brune, rouge, noire, ...). Chaque case de l'histogramme encode la valeur d'une couleur et son nombre de répétition dans la base d'apprentissage. On parle aussi de l'effectif de répétition d'une valeur de couleur. Ainsi l'effectif total représente la somme des valeurs de répétition de toutes les couleurs. Pour déterminer la distribution modélisée par l'histogramme, il suffit de normaliser ce dernier. Une normalisation par l'effectif total permet de définir les fréquences de répétitions pour chaque valeur de couleur. En statistique, la fréquence n'est autre que la probabilité d'apparition d'une valeur. D'où la nomination, table normale de correspondance pour un histogramme normalisé. Ainsi, pour l'image à segmenter en peau, cette table fournira une carte dont chaque case donne la probabilité pour que la valeur du pixel de l'image correspondant soit de la classe peau. Pour déterminer la classe de chaque pixel, sa probabilité qu'il soit de la classe peau est comparée à un seuil qui lui décide de la classe appropriée. Ce modèle est utilisé dans plusieurs travaux de segmentation de la peau tels que [46] [47] [48].

La dimension de l'histogramme dépend en partie de la dimension de l'espace de couleur utilisé. On distingue deux types des Tables normales de correspondances les 3D et les 2D. Ces dernières sont les plus souvent utilisés, tels que [49] qui utilise les composantes  $r$  et  $v$  de l'espace de couleur RGB normalisé, et [50] utilisant la teinte et la saturation ( $H, S$ ). Quant aux auteurs de [51] et [52], ils ont opté pour l'utilisation de l'espace colorimétriques  $RVB$  et donc pour des histogrammes 3D.

Le coefficient de normalisation de l'histogramme peau pour obtenir la table normale de correspondance peut être déterminé, comme précédent, par l'effectif total des pixels de la peau de la base d'apprentissage tel est le cas dans [51], ou encore par le plus grand effectif des différentes valeurs des pixels de la peau comme a été présenté dans [53].

### 1.3.1.2 Classifieur bayésien couplé à l'histogramme

Cette méthode utilise la modélisation de la couleur de la peau et non peau à travers des histogrammes. Ainsi deux histogrammes sont utilisés. Dans chacun d'entre eux, une case encode une valeur de couleur et son nombre de répétition dans les exemples d'apprentissage de la classe représentée. Chaque histogramme est normalisé par son effectif total et permet de modéliser la distribution de la classe (peau et non peau).

Soient  $p(c|peau)$  et  $p(c|\neg peau)$ , les probabilités conditionnelles des distributions peau et

non peau respectivement ;  $c$  est un vecteur de couleur. Notons,  $N(c)$  l'effectif d'une couleur  $c$  et  $N$  l'effectif total d'un histogramme.

Les probabilités conditionnelles peau et non peau s'écrivent comme suit :

$$p(c|peau) = \frac{N_{peau}(c)}{N_{peau}} \quad (1.1)$$

$$p(c|\neg peau) = \frac{N_{\neg peau}(c)}{N_{\neg peau}} \quad (1.2)$$

En utilisant la formule de Bayes, les probabilités qu'un pixel de couleur  $c$  soit de classes peau ou non peau sont :

$$p(peau|c) = \frac{p(c|peau)p(peau)}{p(c|peau)p(peau) + p(c|\neg peau)p(\neg peau)} \quad (1.3)$$

$$p(\neg peau|c) = \frac{p(c|\neg peau)p(\neg peau)}{p(c|\neg peau)p(\neg peau) + p(c|peau)p(peau)} \quad (1.4)$$

Ces deux probabilités sont comparées par rapport à un seuil  $\theta$  ( $0 < \theta < 1$ ). Le pixel est dit pixel peau si la contrainte suivante est vérifiée :

$$\frac{p(skin|c)}{p(\neg skin|c)} > \theta \quad (1.5)$$

La valeur du seuil  $\theta$  est calculée empiriquement et cela à partir de la courbe ROC (*Receiver Operating Characteristic*).

Cette méthode est simple et rapide et ne nécessite que deux histogrammes pour le calcul des probabilités. Elle est utilisée par plusieurs auteurs [54][55][56][52].

### 1.3.1.3 Réseaux bayesiens

Les réseaux bayesiens font partie de la famille des modèles de graphes probabilistes. Ils ont la particularité de fournir une meilleure évaluation des fonctions conditionnelles de densité. Les noeuds d'un réseau représentent des variables aléatoires alors que les arcs déterminent les corrélations entre ces variables [57].

Les réseaux bayesiens les plus populaires sont le Naïf Bayes et le Tree-Augmented Naïf Bayes. Ces réseaux sont qualifiés de naïf car ils supposent l'indépendance des variables.

La structure d'un Tree-Augmented Naïf Bayes est déterminée par la solution maximisant une fonction de vraisemblance sur toutes les données d'apprentissage.

Les auteurs de [20] ont utilisé un réseau naïf pour la modélisation et la classification des pixels en peau et non peau. Ces auteurs ont proposé une méthode pour apprendre la structure du classifieur naïf bayes via des données étiquetées et non étiquetées.

#### 1.3.1.4 Réseaux perceptron multicouches

Les réseaux perceptron multicouches (MLP ou *Multi Layer Perceptron* en anglais), constituent la famille la plus célèbre des réseaux de neurones pour un traitement supervisé. Ce type de réseau est dans la famille générale des réseaux à "propagation vers l'avant" (*feed-forward neural network* en anglais), c'est à dire que l'information se propage dans un sens unique, des entrées vers les sorties sans aucune rétroaction.

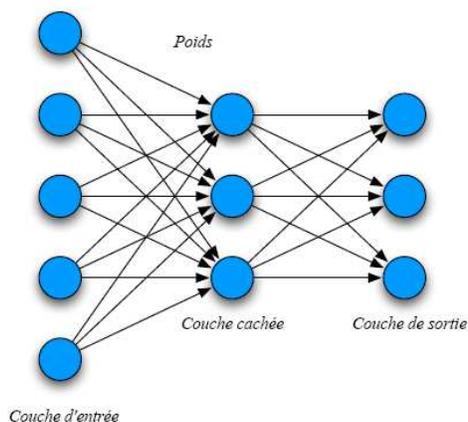


FIGURE 1.5 – Perceptron multicouches.

La Figure 1.5 représente un réseau de perceptron multicouches. Un perceptron multicouches est un réseau comportant  $L$  couches, chaque neurone d'une couche étant totalement connecté aux neurones de la couche suivante. Grâce à sa capacité de classification et de généralisation, le MLP est généralement utilisé dans des problèmes de nature statistique. Or son inconvénient majeur est que ses performances dépendent, parmi d'autres facteurs, du nombre de couches et noeuds cachées utilisées.

Les MLPs ont trouvé leurs applications dans la classification de la peau dans plusieurs espaces de couleur. Dans l'espace  $RGB$ , un réseau perceptron à trois couches est utilisé

par [58][59]. Dans [60] et [61], le MLP modélise la distribution de la peau dans l'espace  $YCrCb$  alors que dans [62] il la modélise dans l'espace de couleur  $CIE - xy$ .

### 1.3.1.5 Cartes auto organisatrices (SOM)

Les cartes auto organisatrices (SOM ou *Self-Organizing Map*) de Teuvo Kohonen ([63], [64]) sont utilisées dans de nombreux domaines d'analyse de données, en particulier dans les espaces de données à grandes dimensions mais elles fonctionnent tout aussi bien pour des données de faibles dimensions. Elles sont aussi appelées algorithme de Kohonen. Il s'agit d'un algorithme stochastique de classement des données, qui intègre une notion de voisinage entre les classes afin de respecter la topologie de l'espace des observations. Il est considéré comme l'un des plus populaires et des plus utilisés des réseaux neuronaux artificiels non supervisés [65]. Les neurones d'un SOM sont organisés dans un plan (chaque neurone est associé à ces voisins) et à chaque neurone est associé un vecteur de poids. La connection entre neurones voisins peut être faite d'une manière rectangulaire ou encore hexagonale comme l'illustre la Figure 1.6.

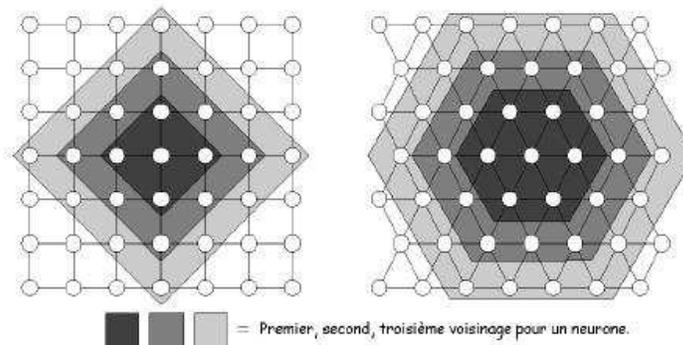


FIGURE 1.6 – Dispositions des neurones du réseau de Kohonen.

Dans [21], l'algorithme SOM est utilisé pour la détection de la peau dans les images comme une première phase dans la détection de visages. Les auteurs ont construits deux SOMs dont l'un modélise la distribution de la couleur de la peau des pixels et l'autre la distribution de la couleur de non peau des pixels. A cette fin, une base d'apprentissage contenant 500 images a été utilisée. Par ailleurs, ces même auteurs ont comparé les performances des SOM et du mélange des gaussiennes dans plusieurs espaces de couleur (rg, HS, TS, ...).

La conclusion de l'étude comparative est que les SOM présentent de meilleurs résultats vis à vis du mélange de gaussiennes indépendamment de l'espace de couleur considéré.

Bien qu'elles soient réputées par leur rapidité, les méthodes non paramétriques exigent un espace mémoire très grand. A leur opposé, les méthodes paramétriques consomment moins d'espace mémoire mais demandent plus de temps pour le calcul de ses paramètres.

### 1.3.2 Méthodes paramétriques

Ce type de méthode permet de modéliser les distributions de la classe peau et la classe non peau, en ajoutant des hypothèses sur les natures de ces distributions. Ainsi, pour les deux distributions, il est demandé de calculer des paramètres de celles-ci (moyenne, variance, ...) et qui dépendent de la méthode utilisée. Nous présentons dans ce mémoire, des extraits de ce groupe de méthodes.

#### 1.3.2.1 Simple gaussienne

Toute distribution d'une classe peut être modélisée par une fonction gaussienne ou par un mélange de gaussienne. En particulier, dans le cas d'une simple gaussienne, la distribution conditionnelle peau s'écrit :

$$p(c|peau) = \frac{1}{2\pi|\Sigma_{peau}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(c - \mu_{peau})^T \Sigma_{peau}^{-1} (c - \mu_{peau})\right) \quad (1.6)$$

où  $c$  est le vecteur couleur. Les paramètres du modèle gaussien sont l'espérance  $\mu_{peau}$  et la matrice de covariance  $\Sigma_{peau}$  définies par :

$$\mu_{peau} = \frac{1}{N_{peau}} \sum_c N_{peau}(c)c \quad (1.7)$$

$$\Sigma_{peau} = \frac{1}{N_{peau} - 1} \sum_c N_{peau}(c)(c - \mu_{peau})(c - \mu_{peau})^T \quad (1.8)$$

et elles sont estimées à partir des exemples constituant la base d'apprentissage.

Cette méthode à été utilisée dans la modélisation de la distribution de la peau et celle de non peau par [66], [67] et [68]. Une fois les distributions conditionnelles peau et non peau sont déterminées, la formule de Bayes appliquée à ces deux distributions permet de calculer la probabilité  $p(peau|c)$ .

### 1.3.2.2 Mélange de gaussiennes

Un modèle plus sophistiqué, capable de décrire des distributions complexe formées, est le modèle gaussien de mélange. Dans la classification de la peau, ce mélange est la généralisation du gaussien simple qui traduit le caractère multimodal de la distribution de couleur de peau.

La probabilité conditionnelle de peau, définie par un mélange de gaussienne s'écrit :

$$p(c|peau) = \sum_{i=1}^k \omega_i p_i(c|peau) \quad \sum_{i=1}^k \omega_i = 1 \quad (1.9)$$

où

- k est le nombre de composants du mélange ;
- $\omega_i$  sont les coefficients du mélange ;
- $p_i$  le noyau d'une gaussienne défini par l'équation (1.6).

Les paramètres à estimer pour une distribution de peau modélisée par un mélange de gaussiennes, sont celles de chaque gaussienne ( $\mu_{peau}$  et  $\Sigma_{peau}$ ) en plus des coefficients  $\omega_i$ .

A cette fin, une variété de l'algorithme EM (*Expectation Maximization*) est utilisée.

La détection de la peau est une problématique ardente à cause de la variation de la couleur de la peau comme a été noté au début de ce chapitre. Par suite, il est plus convenable de modéliser la distribution de la couleur de la peau par un mélange de gaussiennes. Parmi les travaux qui ont adopté l'utilisation d'un mélange gaussien, nous trouvons [69], [70], [71], [72].

### 1.3.2.3 Modèle elliptique de borne

Il s'agit d'un modèle dédié à la modélisation de la distribution de la couleur de la peau, qui a été proposé dans [73]. Une étude comparative sur la base Compaq [52] entre le modèle elliptique de borne, la modélisation par une gaussienne et la modélisation par un mélange de gaussiennes, a montré que le premier modèle présente les meilleurs performances.

La raison derrière une modalisation par une ellipse est justifiée par la forme elliptique que prend la distribution de la couleur de la peau dans n'importe quel espace colorimétriques [73]. Ce modèle est définie par :

$$\Phi(c) = (c - \phi)^T \Sigma_{peau}^{-1} (c - \phi) \quad (1.10)$$

où les paramètres  $\phi$  et  $\Sigma$  sont calculés à partir des exemples d'apprentissage comme suit :

$$\phi = \frac{1}{N_{peau}} \sum_c c \quad (1.11)$$

$$\phi = \frac{1}{N_{peau}} \sum_c N_{peau}(c)(c - \mu_{peau})(c - \mu_{peau})^T \quad (1.12)$$

avec  $N_{peau}$  est le nombre des pixels de peau. L'espérance  $\mu_{peau}$  est calculée par la formule 1.7.

Pour classer un pixel en peau ou non peau, la probabilité  $\Phi(c)$  du vecteur des observations  $c$  est calculée et comparé à un seuil  $\theta$ . Si  $\Phi(c) < \theta$  le pixel sera classé comme pixel de peau, dans le cas opposé il est classé en non peau.

#### 1.3.2.4 Maximum d'entropie

L'entropie est une notion introduite par Shannon en 1945. Elle peut être vue comme une mesure de l'incertitude associée à une distribution de probabilité sur un ensemble d'événements. Le maximum d'entropie est une méthode statistique utilisée pour l'estimation d'une distribution de données. Formellement, le critère de la maximisation de l'entropie sous contraintes, fut développé par Jaynes en 1957. Ce dernier propose de maximiser la mesure d'entropie ou d'information de Shannon sous les contraintes que posent habituellement les probabilités en plus d'autres contraintes données par des informations supplémentaires disponibles ( les exemples d'apprentissages par exemple). Le problème de maximisation de l'entropie nous assure que la distribution de probabilité obtenue est, sous certaines conditions, celle qui est la plus probable parmi l'ensemble de celles qui sont possibles.

La présente méthode est utilisée dans plusieurs domaines pour la détection et la classification des objets. Dans le cadre de la classification de la peau, la maximum d'entropie est utilisé dans [15] pour la modélisation des distributions de couleur peau et non peau. Trois modèles de maximum d'entropie respectant différentes contraintes sur les lois jointes des pixels voisins, sont considérés avec des complexités croissantes : le modèle de référence (baseline), modèle Markov Caché (HMM) et modèle d'arbre au premier ordre (TFOM). Les résultats expérimentaux des comparaisons des trois modèles sur la base Compaq, confirment la supériorité du dernier modèle où les interactions entre les pixels voisins

sont approximées localement par un arbre de Bethe (voir chapitre 2 section 2.3.2) et les paramètres de maximum d'entropie sont estimés par le même arbre.

### 1.3.3 Seuils fixés

L'une des méthodes utilisées pour la détection de la peau, consiste à définir explicitement via un ensemble de règles les frontières qui délimitent le faisceau de la peau dans un certain espace de couleur. Nous citons les deux exemples donnés par [74], [75] et [76], qui analysent le pixel d'une image et décident qu'il s'agit d'un pixel de peau s'il répond aux critères suivants :

– dans le cas d'une luminance uniforme :

$$R > 95, V > 40, B > 20$$

$$\max\{R, V, B\} - \min\{R, V, B\} > 15$$

$$|R - V| > 15, R > V, R > B$$

– dans le cas d'une luminance latérale ou flash :

$$R > 220, V > 210, B > 170$$

$$|R - V| \leq 15, B < R, B > V$$

Les principaux avantages de telles méthodes sont la simplicité de la modélisation et la rapidité en termes d'analyse d'un pixel. Cependant, il est difficile de déterminer les meilleures règles discriminant les pixels de la peau de ceux de non peau dans un espace couleur quelconque, lorsqu'il s'agit d'analyser des images prises sans ou avec peu de contraintes sur les conditions de leurs prises de vue.

## 1.4 Autres modèles pour la distribution de la peau

Bien qu'il est intéressant de se servir de la couleur de la peau comme propriété pertinente pour la localisation des zones de la peau dans une image, quelques chercheurs ont privilégié d'exploiter d'autres caractéristiques pour la segmentation des parties du corps humain dans une image. En particulier, l'extraction des traits caractéristiques ou encore des critères géométriques sont utilisés pour la définition d'un modèle du corps ou

d'une partie de ce dernier. De telles approches sont souvent utilisées aussi bien dans la détection ou la reconnaissance des visages que la détection des mains et de la silhouette d'une personne.

Dans le cas de la détection de visage par exemple, les traits caractéristiques les plus utilisées sont souvent les yeux, la bouche, le nez et le contour extérieur. Chacun de ces traits est associé à une configuration (templates en anglais) connues a priori ou apprises [89]. Dans ce contexte, afin de décider de la présence ou la présence d'un visage dans une image, on fait appel à des contraintes géométriques définissant les positionnements des différentes configurations les unes par rapport aux autres. Aussi, le système proposé par Govindaraju et al. [90], basé sur le contour, modélise le visage comme étant un agencement de trois courbes (sommet, coté droit et coté gauche du visage). La détection de ces courbes et leur regroupement selon leur position relative permet de connaître la position du visage. Parmi les différents types de contour utilisés dans la détection des objets en général, nous citons les contours actifs. L'utilisation de ces derniers dans la détection du visage, permet une adaptation aux variations de forme du visage et conduit ainsi à une meilleure délimitation du visage. De telles méthodes ont été proposées par Waite et al. [77], Craw et al. [78] et Cootes et al. [79]. Une autre approche basée sur les contours consiste en une extraction des contours de l'image et leur mise en adéquation avec une ellipse modélisant un visage [80][81].

Chetverikov [82] utilise non pas des bordures mais des "tâches" et des séquences linéaires de traits de bordure d'orientation similaire. L'étude se base sur un modèle de visage composé de deux tâches sombres pour les yeux, trois tâches claires pour les pommettes et le nez ; les lignes du visage étant caractérisées par les traits d'orientation similaire. La configuration triangulaire des tâches est étudiée ultérieurement.

Dans [83], Yow et Cipolla décrivent un système de détection du visage où les yeux, la bouche et le nez sont extraits par un filtrage en quadrature de phase et sont regroupés selon leur probabilité à constituer un visage. Le problème de l'angle de vue est résolu par l'utilisation d'invariants augmentant la robustesse de leur approche.

Les approches basées sur l'extraction de traits du visage sont plus efficaces que les approches basées sur les contours car il s'agit d'approches ascendantes qui analysent des structures locales de l'image conjointement à des modèles géométriques de visages.

Il existe encore d'autres approches pour délimiter les pixels de la peau dans une image.

Nous citons la détection de mouvement qui est souvent utilisé dans la vidéo. Cette dernière peut être vue comme une succession d'images organisées dans le temps. Ainsi, une des méthodes les plus utilisée pour la détection de la peau, est le suivi du mouvement. Une telle technique offre la possibilité d'orienter la détection de peau vers des zones préférentielles et permet donc une forte réduction de la complexité de la détection et par conséquent une limitation importante des calculs et des traitements [84]. Cependant, les performances de ces systèmes sont fortement réduites lorsque la scène vidéo contient de nombreux objets en mouvement. En outre, dans les systèmes de suivi de mouvement, il est généralement supposé 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. Or, l'information de mouvement n'est jamais utilisée seule pour la détection du fait que l'hypothèse retenue est forte. En effet, cette information est souvent couplée à une information de couleur de peau ([85]) ou encore à un système basé sur la reconnaissance de traits caractéristiques.

## 1.5 Evaluations des modèles existants

Bien entendu pour évaluer différentes méthodes de détection de la peau, il est nécessaire d'envisager le même corpus des images de test. Cependant, dans la littérature les méthodes de la discrimination des pixels de peau et non peau, utilisent des bases de tests différentes et qui sont la plupart du temps inaccessibles. La base Compaq, elle, reste la plus large base des images pour le traitement de la classification de la peau. Cette base contient une des images extraites du Web et qui comprennent une grande diversité des échantillons de la couleur de la peau possible. Cette diversité est due aux différents facteurs agissant sur la définition de la couleur de la peau.

Dans la perspective d'évaluer quantitativement différentes méthodes de détection de classification de la peau, la courbe ROC (*Receiver Operating Characteristic*) est souvent utilisée. Celle-ci permet de combiner les taux des faux positifs et des vrais positifs définies par :

- faux positifs (FP) : proportion des pixels de non peau interprétés comme de peau ;
- vrais positifs (VP) : proportion des pixels de peau correctement interprétés).

Dans un système de détection de la peau où la sortie n'est pas binaire mais probabiliste, l'utilisation d'un seuil (compris entre 0 et 1) permet d'obtenir une image binaire. Des taux d'erreur pour tous les seuils sont récapitulés dans la courbe ROC.

Le tableau suivant (tiré de [41]) synthétise une comparaison entre différentes méthodes de classification de la peau où la base Compaq est utilisée.

TABLE 1.1 – Comparaison entre quelques méthodes de la classification de la peau.

Méthode	Vrais positifs (VP)	Faux positifs (FP)
Modèle bayésien basé sur les histogrammes (RVB)		
[86]	80%	8.5%
[86]	90%	14.2%
[87]	93.4%	19.8%
Simple gaussienne (CrCb)		
[73]	90%	33.3%
Modèle basé sur un mélange de gaussienne (IQ)		
[73]	90%	30%
Modèle basé sur un mélange de gaussienne (RVB)		
[69]	80%	9.5%
[69]	90%	15.5%
SOM en TS		
[88]	78%	32%
Maximum d'entropie en RVB		
[91]	82.9%	10%
Modèle elliptique de borne (CIE-xy)		
[73]	90%	20.9%
Seuils fixés en YIQ		
[87]	94.7%	30.2%

Une méthode est dite performante pour la classification des pixels peau et non peau, si celle-ci présente un taux de vrais positifs élevé pour un taux de faux négatifs minimal. D'après le tableau 1.1, le modèle proposé dans [86], celui du bayésien basé sur les histo-

grammes dans l'espace RVB, offre la meilleure performance. Le modèle proposé dans [91] utilisant le maximum d'entropie dans l'espace RVB, arrive au deuxième rang et les autres modèles le succèdent.

Ainsi, nous pensons qu'il serait judicieux de comparer nos travaux au modèle bayésien basé sur les histogrammes. Cela sera présenté dans le chapitre 5.

## 1.6 Discussion

La détection ou encore la segmentation de la peau consiste à séparer les pixels peau de ceux non peau dans une image. Bien que l'ensemble des méthodes que nous avons présenté dans ce chapitre est non exhaustif, il montre bien l'important intérêt consacré à une telle segmentation dans le domaine de recherche. Les approches de la détection de la peau dans les images couleurs peuvent être subdivisées en deux grandes classes : (1) la détection basée sur l'approche couleur et (2) celle basée sur autre caractéristiques que la couleur.

L'indice couleur est une primitive souvent utilisée pour la détection de la peau. Il est caractérisé par sa simplicité et sa robustesse pour toute transformation géométrique ou changement d'échelle. Dans l'approche couleur, deux axes de recherche sont principalement distingués. Le premier axe, concerne la modélisation de peau/non peau permettant de différencier les pixels de couleur de peau de ceux de non peau. Le deuxième axe concerne le choix de l'espace de couleur adopté pour une meilleure représentation de la couleur de la peau. Dans le contexte du deuxième axe, plusieurs espaces colorimétriques ont été utilisés. Il s'est avéré que l'espace de couleur et la modélisation adaptée sont souvent liés.

Quant à la modélisation de peau/non peau, plusieurs approches aussi bien paramétriques que non paramétriques ou encore à règles fixes sont proposées. Dans le présent chapitre nous avons présenté des exemples de chacune de ces catégories. Les méthodes non paramétriques sont réputées par leur simplicité de calcul mais elles nécessitent un espace mémoire important. A l'opposé, les approches paramétriques ne demandent pas un espace mémoire important mais elles consomment trop en termes de temps de calcul. Aussi, ce type de méthodes impose des contraintes sur la nature des modèles des distributions des deux classes peau et non peau qui peuvent ne pas coïncider avec les distributions réelles de ces deux classes. En ce qui concerne la troisième catégorie, à savoir celle des règles

ou seuils fixes, elle est simple et rapide mais elle n'est pas fiable.

Les techniques basées sur d'autres propriétés que la couleur, sont dédiées à des applications bien précises qui se focalisent sur la détection d'une partie bien particulière du corps ou encore du corps comme une entité globale. Parmi ces méthodes, nous citons celles basées sur l'extraction des traits caractéristiques et de des critères sur la géométrie et celles basées sur la détection de mouvement. Bien que ces deux méthodes aient présenté des résultats encourageants dans la détection de la peau, l'ajout de l'information sur la couleur ne peut que les perfectionner.

La détection de la peau dans les images couleurs demeure une problématique complexe malgré l'importante quantité des recherches menées. Il y a de nombreuses conditions réelles difficiles à prévoir, qui limitent une meilleure modélisation générique de la peau. D'ailleurs les meilleures résultats obtenues dans la de détection de la peau, sont atteintes par des systèmes opérant sous des contraintes limitant les conditions de prise de vue des images. Or, nous nous intéressons à la gestion du Web, plus particulièrement à la détection de la peau sur les images du Web. Ainsi, notre objectif se résume à la construction d'un modèle de prédiction générique permettant de discriminer les pixels de couleur de peau dans une image. pour ce, nous nous appuyons sur des modèles probabilistes graphiques pour la conception d'un modèle de peau discriminant. Ces modèles se caractérisent par leur simplicité, intuitivité et efficacité pour la représentation et la formulation des données. Plus particulièrement, nous proposons d'utiliser des modèles de probabilités indexés par les arbres. Le chapitre qui suit élucide notre propos choisi.



---

## MODÈLES PROBABILISTES INDEXÉS PAR LES ARBRES

Les modèles graphiques probabilistes ont depuis toujours gagné un grand intérêt dans la représentation et la modélisation des connaissances, et cela dans différents domaines d'activités. Cette célébrité est principalement gagnée grâce à l'aspect d'intuitivité et surtout celui de la simplicité dans l'interprétation des interactions entre des données complexes. Les arbres indexant des modèles probabilistes constituent un sous ensemble de ces modèles, souvent utilisés dans les estimations des probabilités.

Le présent chapitre vise à élucider la notion de modèles probabilistes indexés par des arbres. Toutefois, avant d'aborder ce modèle graphique probabiliste, nous rappelons dans une première section les notions de bases qui seront utilisées plus tard dans ce mémoire. La deuxième section tente d'éclaircir la notion de distributions d'arbres et de mettre en valeur l'utilisation des modèles d'arbres pour la modélisation des probabilités. Or, si chaque arbre a sa structure de dépendances, qui le différencie des autres arbres, la question qui nécessite une réponse est, comment est déterminée cette structure ?

Pour répondre à cette question, les sections [2.3](#) et [2.4](#) présentent deux types particulières de structures d'arbres, à savoir les structures définies a priori et les structures déterminées à partir des données. Ainsi, des exemples de chacun des deux types seront décrits.

## 2.1 Modèles graphiques probabilistes : Notions de bases

Les modèles graphiques probabilistes sont dérivés de l'union de deux branches de la mathématique fondamentale, à savoir la théorie des graphes et les modèles probabilistes. Ces deux concepts sont regroupés pour fournir des outils aussi efficaces qu'intuitifs pour représenter une distribution jointe de probabilités sur un ensemble de variables aléatoires. Ils ont révolutionné le développement des systèmes intelligents dans de nombreux domaines. Ces formalismes très puissants permettent des représentations intuitives de la perception humaine concernant la connaissance sur un domaine d'application donné et facilitent la mise en place de modèles performants et clairs. La représentation de la connaissance se base sur la description par des graphes des relations qui existent entre des variables décrivant le domaine d'étude.

Nous notons que les modèles graphiques probabilistes se distinguent en trois catégories : (1) les graphes dirigés ou orientés tels que les réseaux bayésiens, (2) les graphes non orientés tels que les réseaux markoviens et (3) les graphes issues des combinaisons des deux premières catégories. Les distributions des arbres sont considérées comme des cas de figures particuliers des modèles de probabilités des graphes.

Nous rappelons ici au fil de définitions simples, quelques concepts de bases associés aux modèles graphiques et qui sont nécessaires pour les manipulations des distributions d'arbres.

### **Définition 1.** *Grphe*

*Objet mathématique composé de points appelés sommets ou noeuds, et de lignes appelées arêtes connectant un ensemble de ces sommets.*

*Un graphe se définit alors par un triplet  $(N, A, \Psi)$ , où  $N$  est un ensemble de sommets et  $A \subseteq N^2$  est un ensemble de paires de sommets. Quant à  $\Psi$  c'est une fonction définie par – dans le cas d'un graphe orienté :*

$$\Psi : A \rightarrow \{(u, v) \in N \times N\} \quad (2.1)$$

*– dans le cas d'un graphe non orienté :*

$$\Psi : A \rightarrow \{u, v \subseteq N \times N\} \quad (2.2)$$

On désigne par graphe orienté, un graphe dont le passage d'un sommet à un autre est dirigé dans un sens ou/et dans l'autre. Graphiquement ceci est traduit par des flèches reliant les sommets entre eux. Dans ce cas on parle d'arcs reliant les sommets. A l'opposé, dans un graphe non orienté la notion d'orientation d'un sommet à un autre n'existe pas. Ainsi on parle d'arêtes reliant les sommets.

Quand le nombre d'arêtes (ou d'arcs) qui va d'un sommet  $u$  à un sommet  $v$  distinct, ne dépasse pas 1, tous les éléments de  $A$  sont distincts. On note dans ce cas le graphe  $G(N, A)$ . Dans notre étude, nous nous intéressons particulièrement à ce dernier type de graphe. Un exemple d'un tel graphe dans le cas où il est orienté, est illustré par la Figure 2.1.

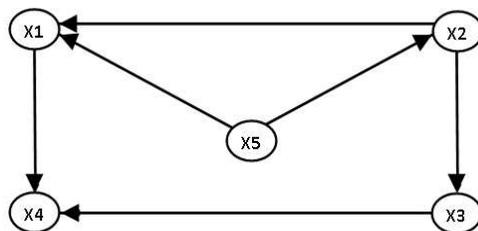


FIGURE 2.1 – Exemple d'un graphe orienté.

Un type particulier de graphe qui est souvent utilisé, est le graphe complet. Celui ci est défini par :

**Définition 2.** *Graphe complet*

*Un graphe est complet si quels que soient deux sommets distincts, il existe un arc (ou une arête) les reliant dans un sens ou dans l'autre.*

Définitions annexes d'un graphe :

- Poids d'une arête : valeur que l'on attribue à une arête.
- Voisinage : deux noeuds  $u$  et  $v$  sont dits si l'arc  $(u, v)$  (ou l'arête) appartient à l'ensemble des  $A$  du graphe.
- Chaîne : suite de sommets reliés par des arêtes.
- Cycle : chaîne dont les extrémités coïncident.
- Connexité : un graphe est connexe si chaque paire de sommets est reliée par une chaîne.

Les arbres ou encore les forêts sont deux cas particuliers d'un graphe. Concernant ces deux modèles, les définitions mathématiques données en théorie des graphes, sont représentées par les définitions 3 et 4.

**Définition 3.** *Forêt*

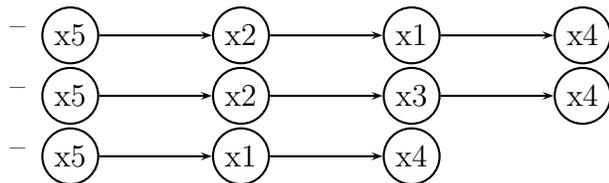
*Un graphe ne contenant aucun cycle (acyclique) est appelé une forêt.*

Le graphe représenté par la Figure 2.1 ne contient aucun cycle. Il s'agit donc d'une forêt illustrée par cette figure.

**Définition 4.** *Arbre*

*Un graphe connexe ne contenant aucun cycle est appelé un arbre.*

Considérons toujours l'exemple du graphe donné à la Figure 2.1, la forêt représentée comporte trois composantes connexes, à savoir :



Chacune de ces composantes constitue un arbre.

A noter qu'un arbre est une forêt connexe. Plusieurs systèmes font recours à ces deux modèles de graphes, principalement grâce au fait qu'ils ne contiennent pas de cycle. Dans la programmation, les cycles créent des problèmes de bouclages ou encore de créations des puits qui peuvent nuire au fonctionnement du système.

Définitions annexes d'un arbre orienté :

– Arc entrant / sortant : soient  $A$  et  $B$  deux noeuds reliés par un arc  $AB$  représenté par



– Racine : noeud de l'arbre qui ne possède aucun arc entrant.

– Feuille : noeud de l'arbre qui ne possède aucun arc sortant.

– Profondeur : maximum des nombres d'arcs reliant la racine aux feuilles.

Dans le contexte de l'optimisation, certaines applications industrielles notamment dans les réseaux de télécommunications cherchent à déterminer des arbres de recouvrements de poids minimal. Un arbre recouvrant d'un graphe est un arbre qui couvre tous les sommets de ce graphe. De plus, le poids d'un arbre couvrant est la somme des poids des arêtes qui constituent l'arbre. Ainsi, un arbre recouvrant minimal est un arbre couvrant d'un graphe dont le poids est minimal.

## 2.2 Distributions d'arbres

Dans un premier temps, nous soulignons que dans la suite de ce mémoire, le mot "arbre" est considéré dans un sens large qui est différent de sa définition (4) en théorie des graphes. Dorénavant, la notion d'arbre correspond à celle de forêt donnée en théorie des graphes et représentée à la définition 3.

Considérons un ensemble de  $n$  variables aléatoires discrètes  $x_i$  constituant un vecteur des observations  $x = (x_1, x_2, \dots, x_n)$ . Les différentes composantes de  $x$  ainsi que les relations entre elles, peuvent être représentées par un graphe  $G(N, A)$ . L'ensemble des noeuds  $N$  est celui des variables discrètes, avec  $|N| = n$  est le cardinal de cet ensemble. La valeur d'un noeuds  $u$  sera notée  $x_u$ ; où  $x_u \in \Omega(u)$  le domaine des valeurs de  $u$ . Quant à  $A$ , il est composé à partir des relations directes entre les variables aléatoires deux à deux. Deux noeuds voisins  $u$  et  $v$  seront noté  $u \sim v$ .

Supposons que le graphe  $G$  est un arbre, noté  $T(N, A)$ , qui possède  $l$  composantes connexes. La relation immédiate qui relie et l'ensemble des noeuds et l'ensemble des arêtes est la suivante :

$$|A| + l = |N| \quad (2.3)$$

Dans le cas particulier où  $T$  est un arbre au sens de la théorie des graphes ( $l = 1$ ), nous retrouvons l'une des propriétés des arbres reliant les ensembles des noeuds et des arcs :  $|A| = |N| - 1$ .

L'arbre  $T$  indexe un modèle de probabilité signifie qu'une probabilité est modélisée par cet arbre; soit  $P(\cdot|T)$  cette probabilité. Considérons les  $P_u$  et  $P_{uv}$  respectivement les probabilités marginales de  $P(\cdot|T)$  au niveau d'un noeud  $u$  et au niveau d'un couple de noeuds  $(u, v)$ , définies comme suit :

$$P_u(x_u) = \sum_{x_v; v \in N; v \neq u} P(x_v|T) \quad (2.4)$$

$$P_{uv}(x_u, x_v) = \sum_{x_s; s \in N; s \neq u, v} P(x_s|T) \quad (2.5)$$

et satisfaisant les deux équations suivantes :

$$P_u(x_u) = \sum_{u_v; v \in N} P_{uv}(x_u, x_v) \quad \forall (u \sim v) \quad (2.6)$$

$$\sum_{x_u \in N} P(x_u | T) = 1 \quad (2.7)$$

Dans [95], les auteurs ont prouvé que la distribution  $P(\cdot | T)$  modélisée par l'arbre  $T$ , dite "distribution de l'arbre  $T$ ", est exprimée par :

$$P(x | T) = \prod_{(u \sim v) \in T} \frac{P_{uv}(x_u, x_v)}{P_u(x_u)P_v(x_v)} \prod_{x_u \in N} P_u(x_u) \quad (2.8)$$

Exemple : La Figure 2.2 représente un arbre non orienté constitué de sept noeuds. La distribution de probabilité modélisée par cet arbre est donnée par la formule suivante :

$$P(x) = \frac{P_{12}(x_1, x_2)P_{32}(x_3, x_2)P_{24}(x_2, x_4)P_{45}(x_4, x_5)P_{65}(x_6, x_5)P_{75}(x_7, x_5)}{P_2^2(x_2)P_4(x_4)P_5^2(x_5)} \quad (2.9)$$

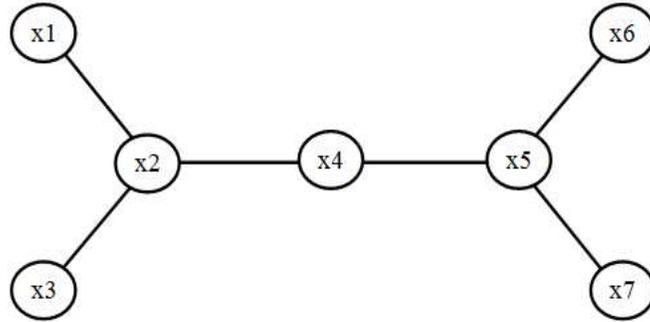


FIGURE 2.2 – Exemple d'un arbre non orienté.

La formule (2.8) traduit le fait que la probabilité modélisée par un arbre est simplement décomposable et déterminée à partir de ses probabilités marginales au niveau des arêtes et de leurs extrémités. Ainsi la structure de l'arbre a une importante influence sur la distribution modélisée. Dans le cas particulier où l'arbre est orienté, la probabilité modélisée par l'arbre  $T$  se résume au produit de probabilités conditionnelles comme suit :

$$P(x | T) = \prod_{u \in N} P(x_u | x_{pa(u)}, T) \quad (2.10)$$

avec  $pa(u)$  désigne le noeud père du noeud  $u$ .

Exemple : La Figure 2.3 illustre un arbre orienté composé de sept noeuds. La distribution de probabilité modélisée par cet arbre est donnée par la formule suivante :

$$P(x) = P_5(x_5) \frac{P_{12}(x_1, x_2)}{P_2(x_2)} \frac{P_{32}(x_3, x_2)}{P_2(x_2)} \frac{P_{24}(x_2, x_4)}{P_4(x_4)} \frac{P_{45}(x_4, x_5)}{P_5(x_5)} \frac{P_{65}(x_6, x_5)}{P_5(x_5)} \frac{P_{75}(x_7, x_5)}{P_5(x_5)} \quad (2.11)$$

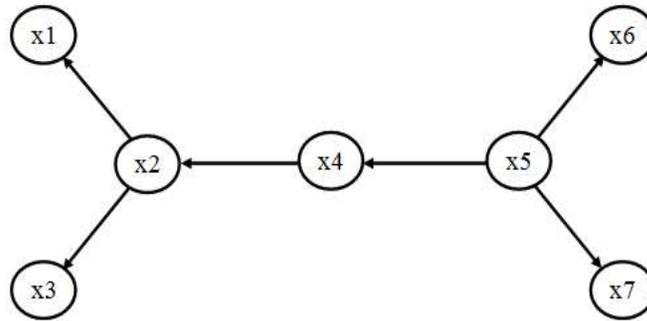


FIGURE 2.3 – Exemple d'un arbre orienté.

Lorsqu'il n'y a pas d'ambiguïté, les désignations "distribution de l'arbre  $T$ " et "arbre  $T$ " seront confondues. De même, les notations  $u$  ou  $x_u$  désigneront le noeud d'un graphe.

Nous attirons l'attention sur le fait que l'arbre  $T$  indexant une probabilité, est souvent nommé arbre de dépendances. Cette appellation découle de la manière dont cet arbre est construit. En effet, l'arbre  $T$  définit et détermine les dépendances (indépendances) entre les éléments du vecteur des observations  $x = (x_1, x_2, \dots, x_n)$  qu'il représente.

Nous notons que les formulations des équations (2.8) et (2.10) sont données respectivement dans le cas d'un arbre non orienté et un arbre orienté. Toutefois, l'équation (2.10) peut être obtenue à partir de l'expression (2.8). Cela est possible en choisissant aléatoirement un ou plusieurs racines de l'arbre non orienté pour définir une relation d'ascendant (père) et de descendant (fils) entre les noeuds de l'arbre. Cette manière de décomposition ou plutôt de séparation de l'arbre total  $T$  en ses sous composantes connexes traduit un aspect d'indépendance entre les sous composantes qui se transmet à la probabilité globale modélisée par un produit des différentes probabilités modélisées par les sous composantes. C'est l'une des propriétés intéressantes des distributions d'arbres qui permet d'avoir des algorithmes d'inférence puissants. Ces derniers font des arbres de dépendances des outils de modélisation et de raisonnement très pratiques.

Revenons aux formules (2.8) et (2.10), et considérons la décomposition de l'arbre que nous venons de décrire. En sachant que

$$P(x_u|x_{pa(u)}, T) = \frac{P_{upa(u)}(x_u, x_{pa(u)})}{P_{pa(u)}(x_{pa(u)})} \quad (2.12)$$

avec  $P(x_u|x_{pa(u)}, T)$  la loi marginal de  $P$  au noeud  $x_u$  qui appartient à l'arbre  $T$ . Nous obtenon donc

$$P(x_u|x_{pa(u)}, T) = \frac{P_{upa(u)}(x_u, x_{pa(u)})}{P_u(x_u)P_{pa(u)}(x_{pa(u)})} P_u(x_u) \quad (2.13)$$

par la suite

$$\prod_{u \in N} P(x_u|x_{pa(u)}, T) = \prod_{(u \sim pa(u)) \in T} \frac{P_{upa(u)}(x_u, x_{pa(u)})}{P_u(x_u)P_{pa(u)}(x_{pa(u)})} \prod_{u \in N} P_u(x_u) \quad (2.14)$$

nous constatons que la deuxième formule (2.10) peut être déduite de la première (2.8).

La structure de l'arbre est un paramètre qui influence et détermine la distribution modélisée. Pour indexer une probabilité, il est donc impératif d'utiliser un arbre dont la structure modélise au mieux cette probabilité. Nous rappelons que dans nos travaux nous nous intéressons à une classification supervisée. Dans ce contexte, les modèles d'arbres existant se divisent en deux catégories : (1) arbres dont les structures sont fournies a priori, que nous appelons aussi arbres à structures prédéfinies et fixes et (2) arbres dont les structures sont déterminées à partir des données. Dans les sections qui suivent, nous discutons ces deux catégories d'arbres.

## 2.3 Arbres à structures prédéfinies et fixes

Dans quelques applications, l'expert intervient pour déterminer l'arbre qui indexe une probabilité. Cette intervention impose une structure de dépendances bien particulière à l'arbre qui doit modéliser les interactions entre les composantes du vecteur des observations représenté. Ci-après, nous présentons deux arbres célèbrement utilisés en imagerie, à savoir l'arbre 4-Star et l'arbre de Cayley.

### 2.3.1 Arbre 4-Star

L'un des arbres les plus utilisés en imagerie est l'arbre 4-Star. Il est souvent appelé pour définir les dépendances entre les voisins d'un pixel dans une image. Ainsi le vecteur des

observations composé et d'un pixel et des voisins de ce pixel, sera représenté par cet arbre. A titre d'exemple, cet arbre est utilisé dans [15] pour estimer les paramètres de la probabilité du maximum d'entropie utilisé pour la détection de la peau.

Un arbre 4-Star  $T_s$  est déterminé à partir de sa profondeur. Ainsi, pour une profondeur égal à 1, l'arbre  $T_s^1$  est déterminé par exemple par un pixel et ses quatre voisins définis dans les quatre directions, comme l'illustre la Figure 2.4.a. L'arbre 4-Star de profondeur égale à 2  $T_s^2$  est construit à partir de l'arbre  $T_s^1$  auquel un noeud est ajouté à chacune de ses feuilles selon leurs directions. L'arbre 4-Star de profondeur 2 est représenté à la Figure 2.4.b. Par récurrence, l'arbre 4-Star de profondeur égale à  $k$  est obtenu en ajoutant un noeud à chacune des quatre feuilles de l'arbre  $T_s^{k-1}$ , et ce dans les quatre directions. Nous remarquons que l'arbre 4-Star doit son appellation au fait que les voisins à considérer sont formés dans les quatre directions à chaque fois.

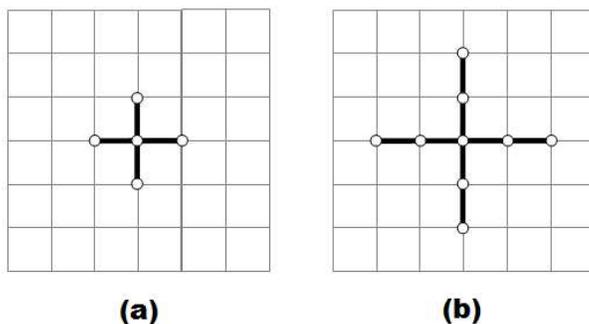


FIGURE 2.4 – Deux arbres 4-Star de profondeurs égales à 1 (a) et à 2 (b).

Un arbre 4-Star est réputé par sa simplicité aussi bien dans la construction de sa structure que dans le calcul de la probabilité qu'il modélise. Or, l'inconvénient majeur d'un tel arbre est que des dépendances exprimées par les voisins proche à la racine mais qui n'appartiennent pas aux quatre directions, sont omises. En plus le nombre de dépendances négligées augmente avec la profondeur de l'arbre.

### 2.3.2 Arbre de Cayley

Un autre arbre dont la structure est fournie a priori est l'arbre de Cayley. Cet arbre apparaît en physique dans le cas de l'approximation, dite de Bethe. Un arbre de Cayley est une structure arborescente avec un taux de branchement  $b$  constant sur toute la

structure à l'exception du centre. Sur un arbre de Cayley, chaque noeud (autre que les feuilles de l'arbre) a  $b + 1$  voisins. La couche  $l$  de l'arbre contient  $(b + 1)b^{l-1}$  noeuds. Cet arbre est caractérisé par son isométrie qui est une propriété importante dans la théorie des graphes.

La Figure 2.5 illustre un exemple de construction d'un arbre de Cayley de profondeur égale à deux et avec un taux de branchement égal à quatre.

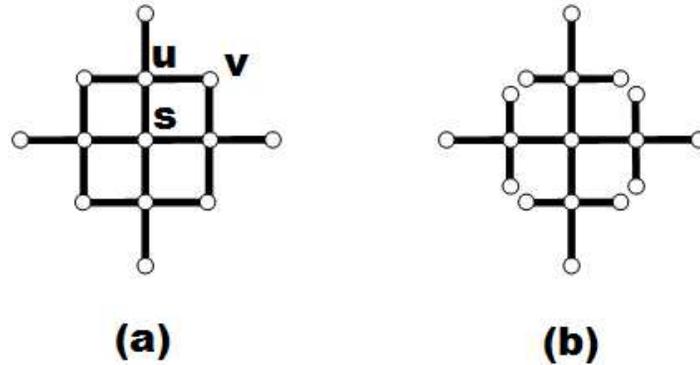


FIGURE 2.5 – Construction d'un arbre de Cayley de profondeur égale à 2. (a) graphe initial. (b) arbre de Cayley déduit.

Etant donné le graphe représenté à la Figure 2.5.(a), le but est de construire l'arbre de Cayley résultant de ce graphe. Pour cela, désignons par  $s$  le centre de l'arbre, à  $s$  nous ajoutons les quatre premiers voisins pour former l'arbre de Cayley de profondeur égale à un et avec un taux de branchement égale à quatre. Ensuite, nous ajoutons à chaque feuille de l'arbre obtenu trois voisins. A cet étape, le noeud  $v$  est éclaté pour éviter la création d'un cycle dans l'arbre ainsi que pour définir un noeud voisin de  $u$ . Le même procédé sera considéré pour la construction du voisin à gauche de  $t$ . Nous remarquons que l'arbre conserve toujours sa propriété d'isométrie. Enfin l'arbre de Cayley de profondeur égale à deux et avec un taux de branchement égal à quatre, qui est donné par la Figure 2.5.(b), est construit.

L'arbre de Cayley est éprouvé présenter de résultats favorables en physique. Cet arbre a été introduit en visions par ordinateur dans [93] comme une alternative pour approximer des estimateurs sur les champs de Markov cachés. Il est aussi utilisé par Zheng dans [15], comme une manière d'approximer la structure de dépendances entre les composantes d'un vecteur des observations déterminé à partir d'un pixel central et ses voisins. L'inconvénient

d'une telle structure d'arbre réside dans l'éclatement de quelques noeuds qui induit des redondances. Ces dernières seront propagées au calcul de probabilité modélisée par l'arbre.

Bien que les arbres dont les structures sont fournies a priori présentent des résultats encourageant dans des cas particuliers, il est préférable de procéder à la définition de la structure d'un arbre à partir des connaissances de l'application étudiée.

## 2.4 Apprentissage de la structure d'un arbre

Lorsque la structure d'un arbre n'est pas fournie a priori, il est possible d'en faire l'apprentissage à partir d'une base de données établie à l'avance. Néanmoins, la détermination de cette structure à partir de données est une problématique NP-difficile pour laquelle de nombreuses méthodes d'apprentissage automatique ont été proposées ces dernières années. L'intérêt de la recherche d'une meilleure structure d'arbre réside dans la quête d'une meilleure approximation de la probabilité jointe des variables aléatoires représentées par les noeuds de cet arbre. Ainsi, nous présentons des travaux utilisant le maximum de vraisemblance pour l'estimation de la probabilité et par suite l'apprentissage de structure d'arbre.

### 2.4.1 Formulation du problème

Soit  $p(\mathbf{x})$  la loi jointe d'un vecteur de  $n$  variables discrètes  $x_1, x_2, \dots, x_n$  représentant le vecteur des observations  $x = (x_1, x_2, \dots, x_n)$ . L'utilisation directe de la loi jointe  $p(\mathbf{x})$  n'est pas toujours facile pour réaliser l'inférence pour deux principales raisons :

La première est due à l'espace mémoire occupé par le vecteur  $x = (x_1, x_2, \dots, x_n)$ , qui dans le cas binaire peut être de l'ordre de  $2^n$ .

La seconde raison est due au temps de traitement de l'inférence dont la complexité est de l'ordre de  $O(2^n)$ .

Pour toutes ces raisons nous avons besoin de simplifier la loi jointe en trouvant les dépendances entre les variables  $x_i$ . Ces dépendances entre variables aléatoires peuvent être représentées sous formes d'un arbre.

Considérons un domaine  $N$  et un ensemble d'observations issu de ce domaine et noté

$D = \{x_1, x_2, \dots, x_n\}$ . Le problème d'apprentissage de structure consiste à définir l'arbre  $T$  qui modélise la distribution de probabilité  $P$  à partir des données  $D$ . Par principe de maximum de vraisemblance, l'estimation de l'arbre  $T$  consiste à chercher le modèle qui maximise la fonction de vraisemblance donnée par :

$$\prod_{x_i \in D} P(x_i, T) \quad (2.15)$$

Nous déduisons la fonction de logarithme de vraisemblance suivant :

$$L(x_1, x_2, \dots, x_n, T) = \sum_{x_i \in D} \log P(x_i, T) \quad (2.16)$$

D'où la meilleure structure de l'arbre  $\tilde{T}$  est la solution exprimée par :

$$\tilde{T} = \arg \max_T \sum_{x_i \in D} \log P(x_i, T) \quad (2.17)$$

Nous attirons l'attention que la probabilité approximée par un arbre  $T$ , sera notée  $P(.|T)$  dans ce mémoire.

## 2.4.2 Résolution du maximum de vraisemblance

Rappelons que la distance entre deux probabilités peut être calculée par une mesure informationnelle dite divergence de Kullbak-Leibler (KL) [94]. Cette mesure est dite aussi entropie croisée ou entropie relative. L'entropie, introduite par Shannon en 1945, est une notion associée à une source qui est modélisée par une variable aléatoire, comme la moyenne de la quantité d'information apportée par les réalisations de cette variable. Elle peut être vue comme une mesure de l'incertitude associée à une distribution de probabilité sur des événements mutuellement exclusifs. Quant à la divergence de Kullbak-Leibler, elle permet la mesure de l'information relative d'une loi de probabilité par rapport à une autre. Cette mesure a aussi été considérée comme une mesure de distance entre ces deux lois, bien qu'elle ne soit pas une vraie distance du fait qu'elle ne vérifie pas l'inégalité triangulaire. La divergence de Kullbak-Leibler entre deux probabilités  $P$  et  $Q$  est exprimée comme suit :

$$KL(P(x), Q(x)) = \sum_{x \in D} P(x) \log \frac{P(x)}{Q(x)} \quad (2.18)$$

avec les convention  $0 \log \frac{0}{Q}$  et  $p \log \frac{p}{0} = \infty$ . Nous remarquons que la distance de Kullback Lebleir  $KL(P, Q) \geq 0$  et égale à 0 si et seulement si  $P(x) = Q(x)$  pour tout  $x$ .

Pour résoudre l'équation du logarithme de vraisemblance donnée en (2.17), Chow Liu ont proposé de minimiser la divergence de KL entre la probabilité réelle et celle lui approximant modélisée par un arbre[92]. Ceci est traduit formellement par :

$$\tilde{T} = \arg \min_T KL(P(x), P(x|T)) \quad (2.19)$$

où  $P$  et  $P(.|T)$  sont respectivement la probabilité réelle et celle modélisée par l'arbre  $T$ . Pour déterminer les caractéristiques de l'arbre optimum  $\tilde{T}$ , nous présentons le calcul développé de l'équation (2.19).

Considérons une distribution de probabilité réelle  $P$  et sa probabilité approximée par l'arbre  $P(.|T)$  dont l'expression est donnée par l'équation (2.8). Nous calculons d'après la définition (2.18) la distance entre  $P(x)$  et  $P(x|T)$  :

$$KL(P(x), P(x|T)) = \sum_{x \in D} P(x) \log \frac{P(x)}{P(x|T)} \quad (2.20)$$

$$= \sum_{x \in D} P(x) \log P(x) - \sum_{x \in D} P(x) \log P_T(x) \quad (2.21)$$

Or,

$$\sum_{x \in D} P(x) \log P(x|T) = \sum_{x \in D} P(x) \log \prod_{(u \sim v) \in T} \frac{P_{uv}(x_u, x_v)}{P_u(x_u)P_v(x_v)} \prod_u P_u(x_u) \quad (2.22)$$

$$= \sum_{x \in D} P(x) \left\{ \sum_{(u \sim v) \in T} \log \frac{P_{uv}(x_u, x_v)}{P_u(x_u)P_v(x_v)} + \sum_u \log P_u(x_u) \right\} \quad (2.23)$$

De plus,

$$\sum_{x \in D} P(x) \sum_u \log P_u(x_u) = \sum_u P_u(x_u) \log P_u(x_u) \quad (2.24)$$

Et

$$\sum_{x \in D} P(x) \log \frac{P_{uv}(x_u, x_v)}{P_u(x_u)P_v(x_v)} = \sum_{(u \sim v) \in T} P_{uv}(x_u, x_v) \log \frac{P_{uv}(x_u, x_v)}{P_u(x_u)P_v(x_v)} \quad (2.25)$$

Il vient que

$$KL(P(x), P(x|T)) = - \sum_{(u \sim v) \in T} KL(P_{uv}(x_u, x_v), P_u(x_u)P_v(x_v)) + \sum_{x_u \in D} H(x_u) - H(x) \quad (2.26)$$

avec  $H$  est l'entropie de Shannon.

Puisque,  $H(x)$  et  $H(x_u)$  sont indépendants de l'arbre, alors minimiser  $KL(P(x), P(x|T))$  revient à maximiser la quantité :

$$\sum_{(u \sim v) \in T} KL(P_{uv}(x_u, x_v), P_u(x_u)P_v(x_v)) \quad (2.27)$$

qui désigne le poids de l'arbre  $T$ .

Nous avons la définition de l'arbre de dépendances de Chow Liu comme suit :

**Définition 5.** *Arbre de Chow Liu*

*C'est l'arbre de dépendances dont le poids donné par*

$$W^T = \sum_{(u \sim v) \in T} KL(P_{uv}(x_u, x_v), P_u(x_u)P_v(x_v)) \quad (2.28)$$

*est maximal.*

D'où découle le théorème 1 de l'arbre de Chow Liu.

**Théorème 1.** .

*Une distribution de l'arbre de dépendances  $P(x|T)$  est une approximation optimum de la distribution  $P(x)$  si et seulement si le poids de l'arbre de dépendances  $T$  est maximum.*

Pour le lecteur intéressé, un exemple détaillé illustrant le fonctionnement de l'algorithme de Chow Liu est décrit dans [92].

Par la suite, plusieurs algorithmes ont été dérivés de l'algorithme de Chow Liu afin de définir la structure d'arbre optimal. Citons à titre d'exemple, Bayes naïf augmenté par arbre (TAN) qui utilise un algorithme modifié de Chow Liu dans sa construction[96]. Le TAN est un type des réseaux bayesiens qui constituent une sous famille des modèles d'arbres. En effet, Ce sont des graphes orientés acycliques qui permettent de représenter et manipuler les indépendances entre les données [95]. Ils étaient principalement créés comme une alternative du classifieur Bayes naïf qui suppose l'indépendance entre toutes les données, condition qui n'est toujours pas vérifiée dans des cas réels.

### 2.4.3 Melange d'arbres

Dans le but d'approximer en mieux une distribution de probabilité réelle, Meila propose de mélanger les probabilités approximées par différents arbres, et ce dans le cadre d'une

classification non supervisée [97]. Le principe de son algorithme consiste à chercher les différents arbres de Chow Liu modélisant des divers aspects d'une distribution de probabilité réelle, et ce sur un ensemble de données. Ainsi, en considérant  $n$  arbres, le modèle du mélange proposé est décrit par la formule :

$$\pi(x) = \sum_{j=1}^m \pi(x|T_p^j)P(T_p^j) \quad (2.29)$$

où  $P$  est une loi sur les arbres avec  $\sum_{j=1}^m P(T_p^j) = 1$ ,  $\pi(x|T_p^j)$  est un champ aléatoire dont les interactions sont décrites par  $T_p^j$  et  $m$  le nombre de composantes mélangées.

La tâche d'un poids  $P(T_p^j)$  particulier est de pondérer la capacité de son modèle correspondant  $T_p^j$  à accomplir sa tâche.

L'estimation des valeurs des paramètres du modèle décrit par l'équation (2.29) revient à trouver les valeurs de ces paramètres qui optimisent la vraisemblance des données d'apprentissage, ou en d'autres termes, de trouver le modèle qui en offrira la meilleur explication de la base d'apprentissage. Dans ce but, Meila utilise une instance de l'algorithme Esperance-Maximisation (en anglais *Expectation Maximization* noté EM). Ce dernier a été proposé par Dempster et al. [100] pour trouver le maximum de vraisemblance des paramètres d'un modèle probabiliste du mélange. Il est partagé en deux étapes successives : l'étape E (pour Espérance) qui consiste à calculer l'espérance de la log-vraisemblance des données complétées conditionnellement aux variables observables et l'étape M (pour Maximisation) dont l'objectif est de maximiser cette espérance afin de mettre à jour les paramètres du modèle.

En posant,  $\theta = (P(T_p^1), P(T_p^2), \dots, P(T_p^m))$  est le paramètre du mélange des probabilités, la fonction de vraisemblance associée au mélange est donnée par

$$L(x, \theta) = \prod_{i=1}^n \pi(x_i) \quad (2.30)$$

En pratique, on maximise le logarithme de la fonction de vraisemblance :

$$\log(L(x, \theta)) = \sum_{i=1}^n \log \left( \sum_{j=1}^m \pi(x_i|T_p^j)P(T_p^j) \right) \quad (2.31)$$

Etnt donnée que la résolution de cette fonction est complexe, cette dernière est complétée par des connaissances sur les ensembles d'arbres considérés. Soit une variable cachée  $z_i^j$  qui

vaut 1 si  $x_i$  appartient à l'ensemble suivant la probabilité  $\pi(x|T_p^j)$  et 0 sinon. L'équation du log-de maximum de vraisemblance à résoudre est :

$$\log(L(x, \theta, z)) = \sum_{i=1}^n \sum_{j=1}^m z_i^j \log(\pi(x_i|T_p^j)P(T_p^j)) \quad (2.32)$$

Par ailleurs les deux étapes classiques de l'algorithme EM qui sont itérées jusqu'à la convergence, sont :

– Etape E : calcul de

$$\Upsilon_j(x^i) = \frac{P(T_p^j) * \pi(x^i|T_p^j)}{\pi(x)} \quad (2.33)$$

– Etape M : détermination du paramètre  $\theta$  par :

$$\Gamma_j = \sum Y_j(x^i) \quad i = 1, 2, \dots, n \quad \text{et} \quad j = 1, 2, \dots, m \quad (2.34)$$

$$P^j(x^i) = \frac{Y_j(x^i)}{\Gamma_j} \quad (2.35)$$

L'instance de l'algorithme EM proposée par Meila est le suivant :

---

**Algorithme 1** Algorithme EM Meila

---

**Entrée :** Base d'apprentissage  $\{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$

$\pi_0 = \{m = n, \pi(x|T_p^j), P(T_p^j)\}$

**Itérer jusqu'à la convergence**

**Etape E :** calculer  $\Upsilon_j^i$  et  $p^j(x^i)$   $j = 1 \dots 3, i = 1 \dots n$

**Etape M :**

$P(T_p^j) \leftarrow \frac{\Gamma_j}{n}$

$\pi(x|T_p^j) \leftarrow$  **Algorithme de Chow Liu pour déterminer  $P_j$**

**Sortie :**  $\pi(x|T_p)$  et  $p(T_p)$

---

L'objectif de cet algorithme consiste à déterminer, pour chacun des aspects différents d'une distribution considérée, les différents arbres de dépendances optimaux donnés par de Chow Liu et les coefficients qui leur correspondent. Les probabilités portées par ces différents arbres sont mélangées pour définir la probabilité du mélange.

## 2.5 Discussion

Le recours aux modèles graphiques probabilistes est motivé par l'attractivité de tels modèles à représenter les connaissances d'une manière intuitive mais aussi efficace. En

effet, la fusion entre la théorie des graphes et les modèles probabilistes, crée un champ fertile pour la description et la manipulation de connaissances complexes.

Les modèles graphiques probabilistes sont partagés en plusieurs catégories. Parmi ces dernières, nous citons les modèles d'arbres indexant des probabilités qui présentent des particularités séduisantes telles que l'absence de cycles et la factorisation de la probabilité modélisée. Dans nos travaux nous nous intéressons à ce dernier modèle de graphes probabilistes. Le problème principal auquel nous nous trouvons confronté, est comment déterminer l'arbre modélisant au mieux les connaissances sur des données? Chaque arbre permet de représenter des relations de dépendances/et indépendances entre les données qu'il modélise. Ces relations permettent de déterminer une structure particulière de l'arbre. Cette structure permet de mettre en valeur les informations importantes pour une meilleure modélisation de la distribution des données. Pour cela, la définition de la structure de l'arbre constitue un axe de recherche doté d'un grand intérêt de la communauté de chercheurs. Les approches proposés dans ce cadre sont partagées en deux ensembles selon la structure de l'arbre considérée : (1) des arbres dont la structure est fournie a priori et (2) des arbres dont la structure est déterminé par les données.

Dans le cas d'arbres à structure fournie a priori, nous avons présenté deux arbres communément utilisés aussi bien en imagerie qu'en d'autres applications, à savoir l'arbre 4-Star et l'arbre de Cayley aussi connue sous le nom d'arbre de Bethe. Chacun de ces deux arbres détermine une structure de dépendances particulière entre les données et qui permet de modéliser de la même manière les distributions de probabilités quelque soit le domaine d'application. Or, c'est une hypothèse forte car en réalité chaque distribution à ses propres caractéristiques. Pour cela, bien que ce genre d'arbres ait présenté des résultats satisfaisant dans des quelques domaines, ce gain ne peut forcément pas être généralisés aux autres domaines.

Pour alléger l'hypothèse d'arbres dont la structure est fournie a priori, il est possible de définir la structure d'un arbre à partir des données. Un apprentissage sur les données exploite les connaissances sur les différents sur celles-ci et définit les relations d'interactions pertinentes entre elles. Ces interactions entre les noeuds d'un arbre sont traduites par des probabilités marginales calculées au niveau d'une arête et ses extrémités. Le problème d'apprentissage de la structure d'un arbre est présenté dans ce chapitre comme une fonction de maximum de vraisemblance à résoudre. Une manière de résoudre cette dernière a

été fournie par Chow Liu. Ces chercheurs ont présenté une heuristique pour déterminer la meilleure structure d'arbre modélisant une probabilité conditionnelle. Des méthodes dérivées de cette approche ont été présentées par la suite. En particulier, l'algorithme de Meila exploite dans le cas d'une classification non supervisée, la notion de variable cachée dans un réseau pour proposer un mélange des probabilités modélisées sur un ensemble d'arbres. Chacun de ces arbres considérés représente une caractéristique particulière sur des données. Dans ces deux algorithmes de base, l'arbre modélise une distribution de probabilité à partir des exemples sur celle-ci. Ainsi, la structure de l'arbre fournie par l'apprentissage ne tient compte que du comportement de la distribution de la probabilité qu'il modélise.

L'objectif du chapitre suivant consiste à améliorer les modèles de distributions d'arbres présentés ici.

Le présent chapitre présente des nouvelles distributions d'arbres, à la fois modélisant des distributions de probabilités et approximant au mieux les probabilités réelles. En fait, la nécessité d'un nouveau modèle est exigée afin de remédier aussi bien au problème des arbres dont les structures sont fournies à priori qu'à celui des arbres dont la définition des structures ne considère que des exigences sur les distributions de probabilités qu'elles indexent. D'une part, l'utilisation des arbres dont les structures sont fournies à priori, impose et met en valeur certaines interactions entre les données représentées. Ces interactions ne reflètent forcément pas les vraies dépendances (indépendances) entre les données et par suite les probabilités modélisées seront erronées. D'autre part, les modèles d'arbres indexant une probabilité indépendamment de son milieu d'étude, ne reflètent toujours pas les contraintes réelles. En effet, entre deux distributions de données ou de classes, des chevauchements ou des interactions existent souvent. Cette intersection non vide entre les classes remet en question les deux modèles d'arbres cités auparavant.

Le but de ce chapitre est de définir notre vision pour les distributions d'arbres, définie en particulier dans le cadre de la détection de la peau et qui permet de remédier aux deux problèmes soulignés ci-dessus. Sachant que notre modèle est fondé sur un critère d'optimisation, notre travail consiste à déterminer différents critères possibles dans le cadre de la détection de la peau. Ainsi, nous présentons au fil d'eau, un ensemble de modèle basé sur les arbres en vue d'aboutir à une meilleure modélisation des distributions peau et non peau.

### 3.1 Quelques notations

Dans la présente section, nous fixons quelques notations qui faciliteront la compréhension de la suite de ce mémoire.

Notons  $s$  un pixel et  $x_s$  sa valeur. Le couple  $(i_s, j_s)$  désignera les coordonnées de  $s$  dans une image. La plupart des approches de détection de la peau fait recourt au classement d'un pixel indépendamment de ceux qui l'entoure. Cependant, étant donné que la texture de la peau est homogène, une corrélations entre le pixel à classifier et ses voisins existe. Nous incorporons la notion de voisinage dans la classification de chaque pixel. Ainsi, la classe d'un pixel n'est autre que la classe de voisinage considéré pour ce pixel.

Nous définissons le système de voisinage d'un pixel  $s$  comme suit :

$$V_s^r = \{(i, j) / |i - i_s| < r, |j - j_s| < r\} \setminus \{(i_s, j_s)\} \tag{3.1}$$

où  $r$  est un entier positif.

La Figure (3.1) illustre trois exemples de voisinages. Dans le premier voisinage représenté en (a), la valeur de  $r$  est égale à 1. Dans le second (b),  $r$  prend la valeur 3, la taille de l'ensemble des voisins est donc égale à 8. Finalement, le voisinage donné par (c) est obtenu pour une valeur de  $r$  égale à 5, d'où le cardinal de l'ensemble des voisins est égal à 24.

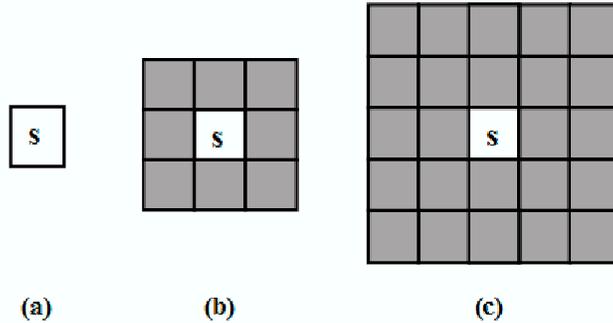


FIGURE 3.1 – Exemples de systèmes de voisinages d'ordre  $r$  (avec  $r = 0, 2, 4$ ).

Considérons maintenant le vecteur  $(x_1, x_2, \dots, x_n)$  issu d'un bloc de pixel composé d'un pixel et ses voisins; avec  $n = r^2$ . La taille du bloc considérée dans l'évaluation des algorithmes proposés dans cette thèse ainsi que l'espace de couleurs utilisé, seront spécifiés dans le protocole expérimental présenté au chapitre 5.

## 3.2 Un nouveau modèle d'arbres pour indexer des probabilités

### 3.2.1 Principe de base et motivation

Soit un vecteur des observations  $x = (x_1, x_2, \dots, x_n)$ . Notons respectivement  $C_1$  et  $C_2$  les deux classes peau et non peau étiquetées par '1' et '2', et représentées par la variable aléatoire  $y$  qui prend des valeurs dans l'ensemble  $\{1, 2\}$ . Nous utilisons des arbres pour indexer les deux probabilités conditionnelles  $P(y_s = 1|x)$  et  $P(y_s = 2|x)$ ; avec  $P(y_s = C_i|x)$  la probabilité mesurer l'incertitude que le pixel  $s$  dont le vecteur des observations est  $x$ , soit appartenant à la classe  $C_i$  ( $i = 1, 2$ ). Pour cela, il existe deux solutions pour procéder à l'indexation de chaque distribution de probabilité : utiliser un arbre dont la structure est fournie à priori, ou encore définir et apprendre la structure d'arbre qui modélise la distribution de probabilité. Du moment que la définition de la structure de l'arbre à partir des données reflète les indépendances entre celles-ci, il semble judicieux d'opter pour la deuxième solution. A cette fin, il suffit de considérer une base d'apprentissage partagée en deux sous bases : une première ne contenant que des exemples de la classe  $C_1$  (des échantillons de la peau) et une deuxième ne contenant que des exemples de  $C_2$  (aucun échantillon de la peau). Chaque sous base permet d'apprendre la structure de l'arbre indexant la probabilité de sa classe. On voit bien que chacune des structures d'arbres est déterminée indépendamment de l'autre. Autrement dit, dans l'apprentissage de la structure de l'arbre d'une classe, les caractéristiques de l'autre classe ne sont pas prises en considération.

La Figure 3.2 illustre le principe de construction des arbres indexant les probabilités peau et non peau à partir des bases des exemples des deux classes en question. La question qui se pose est : est ce que l'indépendance dans les constructions des deux arbres ne va-t-elle pas augmenter la marge des éléments classés avec ambiguïtés ?

Nous savons (comme a été souligné dans l'introduction du document) que la classification de la peau dans les images couleurs est une problématique ardente. La difficulté de la classification des pixels des images par rapport à ces deux classes (peau et non peau) réside principalement dans deux facteurs : (1) la variation de la couleur de la peau sous différentes conditions (peau blanche, peau noire, peau jaune, ...) et (2) la multitude des objets qui

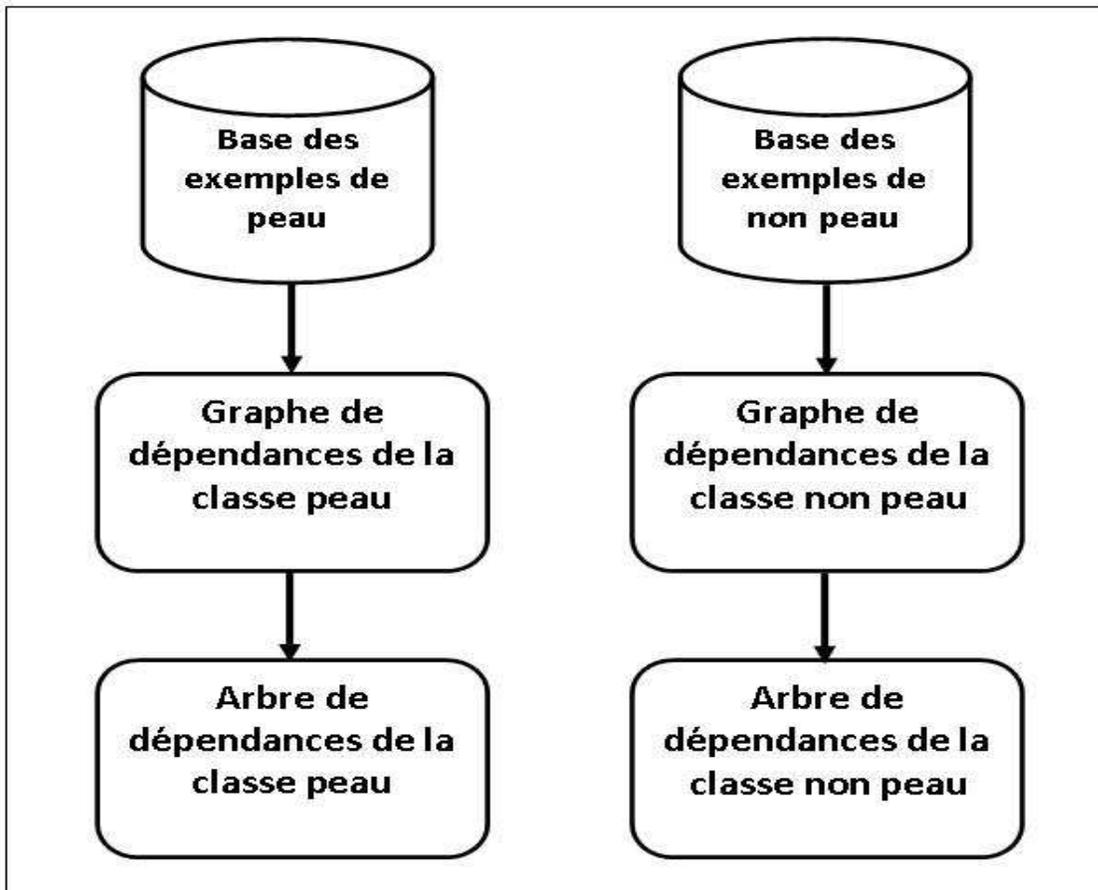


FIGURE 3.2 – Principe de construction des arbres indexant les probabilités peau et non peau.

ont la même couleur que celle de la peau humaine (bois, sable, charbon, ...). Ces deux facteurs sont traduits principalement par le problème d'interclasse entre la classe peau et la classe non peau, qui est le noyau de la problématique de la détection de la peau. Il paraît donc nécessaire de gérer le problème d'interclasse entre les deux distributions  $P(y_s = 1|x)$  et  $P(y_s = 2|x)$  lors de la construction des deux arbres indexant ces probabilités. Or, les deux arbres construits précédemment pour indexer ces deux probabilités conditionnelles, ne prennent pas en considération ce problème.

En tenant compte de ce constat, nous proposons par la suite une nouvelle solution pour indexer les deux probabilités peau et non peau, en utilisant les arbres.

### 3.2.2 Modèle des arbres proposé

Les distributions des arbres permettant de bien classer les pixels des images en peau et non peau, ne doivent pas ignorer lors de leurs constructions le comportement des deux classes l'une par rapport à l'autre. Dans ce contexte, nous proposons une nouvelle distribution d'arbres qui permet d'exploiter les connaissances sur les deux classes pour une meilleure classification de la peau. Nous nommons cette proposition "arbre de dépendances à bi-probabilité peau et non peau" que nous définissons comme suit :

**Définition 6.** *Arbre de dépendances à bi-probabilité peau et non peau*

*C'est l'arbre dont l'ensemble des noeuds est constitué des composantes du vecteur d'observations  $x$  et qui indexe avec performance les distributions de probabilités des deux classes peau et non peau en même temps.*

A noter qu'ici, le mot performance signifie que les deux distributions de probabilités modélisées par l'Arbre de dépendances à bi-probabilité peau et non peau, sont les optimales pour les deux classes. Autrement dit, ces distributions sont les plus proches aux probabilités réelles de peau et non peau.

Dans le présent mémoire nous nous intéressons à l'indexation des probabilités peau et non peau. Ainsi, par abus de langage, nous acceptons que "l'arbre de dépendances à bi-probabilité peau et non peau" soit noté "arbre de dépendances à bi-probabilité".

Le choix d'un arbre unique pour l'indexation des deux probabilités conditionnelles  $P(y_s = 1|x)$  et  $P(y_s = 2|x)$  est principalement déduit de la contrainte de complémentarité des deux classes peau et non peau. Par ailleurs, cette structure de dépendance particulière entre les éléments du vecteur d'observations permettra de gérer au mieux les interclasses entre les deux classes considérées. La question qui se maintient maintenant est la suivante : comment déterminer l'arbre de dépendances à bi-probabilité indexant conjointement les deux distributions peau et non peau ?

Afin de répondre à la question posée, nous considérons un graphe complet  $G(N, A)$  tel que l'ensemble des noeuds ( $N$ ) de  $G$  est constitué à partir des composantes du vecteur des observations  $x = (x_1, x_2, \dots, x_n)$  considéré, alors que l'ensemble des arcs  $A$  est composé des relations entre tous les noeuds, deux à deux. Etant donné que l'arbre de recouvrement optimal (maximal ou minimal) a prouvé son efficacité dans plusieurs applications appartenant à divers secteurs d'activités, nous choisissons que l'arbre de dépendances à

bi-probabilité soit l'arbre de recouvrement optimal du graphe complet  $G$ . Toutefois, l'optimisation doit être calculée à partir d'un critère d'optimisation. La difficulté qui persiste est le choix de ce critère qui permet d'estimer au mieux les probabilités conditionnelles réelles  $P(y_s = 1|x)$  et  $P(y_s = 2|x)$ .

La Figure 3.3 illustre le principe de construction d'un arbre de dépendances à bi-probabilité à partir des bases des exemples des deux classes peau et non peau.

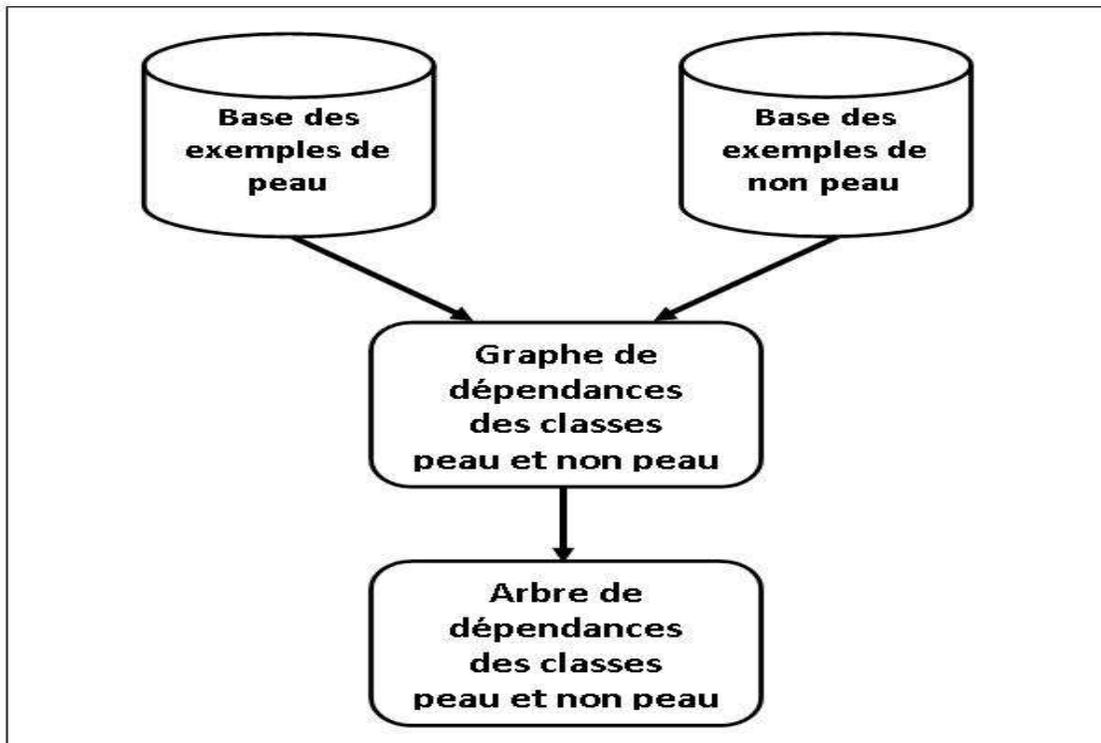


FIGURE 3.3 – Principe de construction de l'arbre de dépendance à bi-probabilité peau et non peau.

La définition du critère d'optimisation du nouveau modèle d'arbres, fera l'objet des sections qui succèdent en vue de déterminer celui le plus discriminant pour la classification des pixels peau et non peau dans une image.

Nous précisons que les notations présentées dans ce paragraphe seront réutilisées dans la suite du chapitre. Aussi, afin de simplifier les notations nous posons :

- $P(y_s = 1|x) = p(x)$  ;
- $P(y_s = 2|x) = q(x)$  ;
- $P(x|T)$  désigne la probabilité  $P$  modélisée par l'arbre  $T$  ;

- $u$  et  $x_u$  sont confondues. Elles désignent un noeud du graphe  $G$  ;
- $P_u$  est la loi marginale du noeud  $u$  ;
- $P_{uv}$  est la loi marginale du couple de noeuds  $(u, v)$ .

### 3.3 Arbres de dépendances à bi-probabilité peau et non peau Simple

#### 3.3.1 Critère d'optimisation

Certes, il est clair que pour définir l'arbre de dépendances à bi-probabilité modélisant au mieux les deux classes peau et non peau, il faut préciser son critère d'optimisation. Or, déterminer ce dernier revient à trouver une fonction caractéristique  $\phi$  de l'arbre, décrivant les relations entre les distributions de probabilités modélisées par cet arbre. Ensuite, grâce à un estimateur de maximum de vraisemblance fondé sur  $\phi$ , nous pouvons définir l'arbre en question. Mais, comment la fonction caractéristique  $\phi$  peut-elle être formalisée ?

Dans un premier temps, nous pouvons nous inspirer des arbres de Chow Liu appliqués à la détection de la peau, pour définir l'arbre de dépendances à bi-probabilité. Autrement dit, au lieu de définir un arbre de Chow Liu pour une classe, une émergence des deux arbres indexant les distributions peau et non peau peut former l'arbre de dépendance à bi-probabilité. Pour cela, nous considérons les critères de Chow Liu portant sur les deux classes en même temps. Vu que la divergence de Kullback - Leibler est non-négative, une fonction caractéristique de l'arbre de dépendances à bi-probabilité peut être définie par :

$$\phi_1(T) = KL(p(x), p(x|T)) + KL(q(x), q(x|T)) \quad \forall T \subset G \quad (3.2)$$

Par ailleurs, le critère d'optimisation de l'arbre de dépendances à bi-probabilité peau et non peau, noté  $\tilde{T}$ , est formalisé comme suit :

$$\zeta = \min_{T \subset G} \phi_1(T) \quad (3.3)$$

Par conséquent, nous présentons la définition de l'arbre de dépendances à bi-probabilité que nous nommons " arbre de dépendances à bi-probabilité peau et non peau Simple" :

**Définition 7.** *Arbre de dépendances à bi-probabilité peau et non peau Simple*

*C'est l'arbre de dépendances à bi-probabilité  $\tilde{T}$  minimisant la somme des divergences de Kullback-Leibler entre les probabilités réelles et celles modélisées des deux classes peau et non peau. Formellement, l'arbre  $\tilde{T}$  est une solution du critère d'optimisation, exprimé par :*

$$\tilde{T} = \arg \min_{T \subset G} \{KL(p(x), p(x|T)) + KL(q(x), q(x|T))\} \quad (3.4)$$

avec  $p(x|T)$  et  $q(x|T)$  sont les probabilités modélisées par l'arbre  $T$ .

L'arbre donné par la définition 7 peut être vu comme celui représentant un compromis entre, d'un côté la distance entre la probabilité peau réelle et celle estimée et de l'autre côté la distance entre la probabilité non peau réelle et celle estimée.

Dans le paragraphe qui suit, nous développons l'équation (3.4) en vue de définir les caractéristiques de l'arbre de dépendances à bi-probabilité peau et non peau Simple.

### 3.3.2 Modèle de l'arbre de dépendances à bi-probabilité peau et non peau Simple

Supposons pour le moment que l'arbre de dépendances à bi-probabilité peau et non peau Simple existe. Dans ce cas, l'équation (3.4) admet une solution dans le graphe  $G$ . Or, quelles sont les pondérations des arcs du graphe  $G$ ? Pour répondre à cette question nous proposons de développer l'équation (3.4). Suite aux calculs présentés dans le chapitre précédent section (2.4.2), nous avons :

$$KL(P(x), P(x|T)) = - \sum_{(u \sim v) \in T} KL(P_{uv}(x_u, x_v), P_u(x_u)P_v(x_v)) + \sum_{u \in N} H(x_u) - H(x) \quad (3.5)$$

avec  $H$  est l'entropie de Shannon.

Par substitution de la probabilité  $P$  une fois par  $p$  et une autre par  $q$ , nous déduisons que la quantité  $KL(p(x), p(x|T)) + KL(q(x), q(x|T))$  s'écrit sous la forme :

$$KL(p(x), p(x|T)) + KL(q(x), q(x|T)) = \overbrace{- \sum_{(u \sim v) \in T} KL(p_{uv}(x_u, x_v), p_u(x_u)p_v(x_v))}^A$$

$$+ \overbrace{\sum_{x \in N} q(x) \log q(x) - \sum_{u \in N} q(x_u) \log q(x_u) + \sum_{x \in N} p(x) \log p(x) - \sum_{u \in N} p(x_u) \log p(x_u)}^B$$

$$- \underbrace{\sum_{(u \sim v) \in T} KL(q_{uv}(x_u, x_v), q_u(x_u)q_v(x_v))}_C \quad (3.6)$$

Etant donnée que le terme B est indépendant de l'arbre  $T$ , résoudre l'équation (3.4) revient à maximiser les quantités A et C qui forment :

$$W^T = \sum_{(u \sim v) \in T} KL(p_{uv}(x_u, x_v), p_u(x_u)p_v(x_v)) + KL(q_{uv}(x_u, x_v), q_u(x_u)q_v(x_v)) \quad (3.7)$$

Cette quantité désigne simplement le poids de l'arbre recouvrant  $T$ . Nous déduisons que les pondérations du graphe  $G$  sont données par :

$$KL(p_{uv}(x_u, x_v), p_u(x_u)p_v(x_v)) + KL(q_{uv}(x_u, x_v), q_u(x_u)q_v(x_v)) \quad (u \sim v) \in A \quad (3.8)$$

En conclusion, le modèle de l'arbre de dépendances à bi-probabilité peau et non peau Simple consiste à déterminer l'arbre recouvrant maximal d'un graphe dont les pondérations sont exprimées par l'équation (3.8). Celles-ci formalisent les degrés d'indépendances de deux noeuds voisins pour les deux distributions peau et non peau.

La Figure 3.4 illustre quelques entrées (images couleurs) et sorties (images aux niveaux de gris) de deux détecteurs de la peau. Les images sur la deuxième et la troisième colonne représentent respectivement les sorties du détecteur fondé sur les arbres de Chow Liu et celui basé sur l'arbre de dépendances à bi-probabilité peau et non peau Simple.

Les exemples des images couleurs utilisées prennent en considération quelques variations de prise de vue et la diversité de la couleur de la peau. Les différences de la détection de la peau entre les sorties représentées sur la deuxième colonne et celles de la troisième colonne, peuvent être notées autour de la bouche sur la première image en haut, le front et la barbe sur la deuxième image et finalement sur l'arrière plan complexe et la bouche de la troisième image. Bien que visuellement la qualité de la détection de la peau en utilisant les deux modèles comparés ne soit pas très remarquable, nous notons que le modèle de l'arbre de dépendance à bi-probabilité peau et non peau Simple permet d'améliorer la détection de la peau par rapport au modèle de l'arbre de Chow Liu. Toutefois, des erreurs de classifications sont à noter encore sur les sorties du nouveau détecteur.

Le modèle de l'arbre de dépendances à bi-probabilité Simple prend en considération le comportement des probabilités modélisées vis-à-vis de leurs distributions réelles. En effet, son critère d'optimisation est traduit par la minimisation des distances entre les



FIGURE 3.4 – Quelques entrées et sorties de deux détecteurs de peau. Les images de la deuxième et troisième colonne correspondent respectivement à celles obtenues par le détecteur fondé sur l'arbre de Chow Liu et celui à base de l'arbre de dépendances à bi-probabilité peau et non peau Simple.

probabilités réelles et celles modélisées de part et d'autre des deux classes peau et non peau. Par ailleurs, par le biais de l'utilisation d'un seul graphe de dépendances et pour la classe peau et pour la classe non peau, le problème d'interclasse entre ces dernières est implicitement tenu en compte. Ainsi chaque arbre de dépendances à bi-probabilité du graphe considéré, cache une notion d'interactions entre les deux classes peau et non peau, en particulier l'arbre dit arbre de dépendances à bi-probabilité Simple. Or, la question concernant la manière dont ces interactions sont gérées, reste formellement sans réponse. D'où la nécessité de définir un autre modèle d'arbres de dépendances à bi-probabilité.

## 3.4 Pour un Meilleur Arbre Recouvrant

### 3.4.1 Critère d'optimisation

Étant donné que le noyau de la problématique de la détection de la peau est l'interclasse entre les classes peau et non peau, il semble judicieux que le critère d'optimisation le plus approprié à utiliser doit exploiter la notion d'interclasse. Ceci permet de minimiser les interactions entre les deux classes et d'estimer au mieux les probabilités peau et non

peau réelles. La limitation du modèle de l'arbre de dépendances à bi-probabilité Simple est causée par l'ignorance de la gestion formelle du problème d'interclasse, il est donc préférable que le critère d'optimisation du nouveau modèle remédie à cet handicap.

Lorsque nous évoquons le problème d'interclasse entre deux classes, nous faisons automatiquement référence à la similarité entre ces deux classes. Toutefois, pour comprendre la similarité il faut définir une métrique adaptée. Nous choisissons une métrique informationnelle, la divergence de Kullback Leibler qui a une relation étroite avec le gain d'information, comme moyen pour mesurer la similarité entre les deux distributions de probabilités peau et non peau modélisées.

Il est intuitif que les probabilités modélisées par l'arbre de dépendances à bi-probabilité  $T$ , estiment au mieux les probabilités réelles lorsque la mesure de similarité entre ces dernières est aussi grande que possible. De là, la divergence de Kullback-Leibler entre les distributions de probabilités modélisées doit être maximale. Ainsi, la fonction caractéristique  $\phi_2$  de l'arbre  $T$  spécifiant les probabilités modélisées et estimant les probabilités conditionnelles réelles  $P(y_s = 1|x)$  et  $P(y_s = 2|x)$  peut être définie comme suit :

$$\phi_2(T) = KL(p(x|T), q(x|T)) \quad \forall T \subset G \quad (3.9)$$

avec  $p(\cdot|T)$  et  $q(\cdot|T)$  sont respectivement les probabilités modélisées par  $T$  et approximant les probabilités réelles  $p$  et  $q$ .

Le critère d'optimisation de l'arbre de dépendances à bi-probabilité peau et non peau, se traduit par la formule suivante :

$$\zeta = \max_{T \subset G} \phi_2(T) \quad (3.10)$$

Par conséquent, nous présentons la définition 8 qui résume le deuxième modèle de l'arbre de dépendances à bi-probabilité peau et non peau défini.

**Définition 8.** *Meilleur Arbre Recouvrant*

*C'est l'arbre de dépendances à bi-probabilité peau et non peau  $\tilde{T}$  maximisant la divergence de Kullback-Leibler entre les probabilités peau et non peau modélisées. Formellement, l'arbre  $\tilde{T}$  est la solution de l'équation du critère d'optimisation, qui s'écrit comme suit :*

$$\tilde{T} = \arg \max_T \{KL(p(x|T), q(x|T))\} \quad (3.11)$$

*avec  $p(\cdot|T)$  et  $q(\cdot|T)$  les probabilités approximées par l'arbre  $T$ .*

Dans la suite, nous proposons de développer l'équation (3.11) en exploitant la définition de la divergence de Kullback-Leibler, dans l'objectif de déterminer les caractéristiques du Meilleur Arbre Recouvrant.

### 3.4.2 Modèle du Meilleur Arbre Recouvrant

La présente sous section a un double objectif : il s'agit, d'une part, de définir les pondérations du graphe  $G$  représentant le vecteur des observations  $x$  et d'autre part, de déterminer les caractéristiques du Meilleur Arbre Recouvrant. Pour ce, il est certain qu'il est indispensable d'exploiter l'équation (3.11), que le Meilleur Arbre Recouvrant satisfait.

Par hypothèse, nous supposons qu'il existe un arbre  $T$  dont les noeuds sont les  $n$  composantes du vecteur des observations  $x = (x_1, x_2, \dots, x_n)$  et qui modélise les distributions de probabilités approximées  $p(x|T)$  et  $q(x|T)$ . Par définition du Meilleur Arbre Recouvrant, les distributions de probabilités peau et non peau modélisées par cet arbre sont estimées par résolution de l'équation (3.10) qui peut aussi s'écrire sous la forme :

$$\begin{aligned} KL(p(x|T), q(x|T)) &= \sum_{x \in V} p(x|T) \log \frac{p(x|T)}{q(x|T)} \\ &= \sum_{x \in N} p(x|T) \log p(x|T) - \sum_{x \in N} p(x|T) \log q(x|T) \end{aligned} \quad (3.12)$$

D'une part, nous avons :

$$\begin{aligned} \sum_{x \in N} p(x|T) \log p(x|T) &= \sum_{x \in N} p(x|T) \sum_{x_u \in N} \log p_u(x_u) \\ &+ \sum_{x \in N} p(x|T) \sum_{(u \sim v) \in T} (\log p_{uv}(x_u, x_v) - \log p_u(x_u)p_v(x_v)) \end{aligned} \quad (3.13)$$

et d'autre part, nous avons :

$$\begin{aligned} \sum_{x \in N} p(x|T) \log q(x|T) &= \sum_{x \in N} p(x|T) \sum_{x_u \in N} \log q_u(x_u) \\ &+ \sum_{x \in N} p(x|T) \sum_{(u \sim v) \in T} (\log q_{uv}(x_u, x_v) - \log q_u(x_u)q_v(x_v)) \end{aligned} \quad (3.14)$$

Par soustraction de (3.13) et (3.14), l'équation (3.12) devient :

$$KL(p(x|T), q(x|T)) = \sum_{x \in N} p(x|T) \sum_{x_u \in N} \log \frac{p_u(x_u)}{q_u(x_u)}$$

$$+ \sum_{x \in N} p(x|T) \sum_{(u \sim v) \in T} \left( \log \frac{p_{uv}(x_u, x_v)}{q_{uv}(x_u, x_v)} - \log \frac{p_u(x_u)p_v(x_v)}{q_u(x_u)q_v(x_v)} \right) \quad (3.15)$$

Nous déduisons alors que :

$$KL(p(x|T), q(x|T)) = \sum_{x_u \in N} KL(p_u(x_u), q_u(x_u)) + \sum_{(u \sim v) \in T} \{KL(p_{uv}(x_u, x_v), q_{uv}(x_u, x_v)) - KL(p_u(x_u), q_u(x_u)) - KL(p_v(x_v), q_v(x_v))\} \quad (3.16)$$

Il est à noter que pour tout  $x_u \in N$ , la mesure de la quantité  $KL(p_u(x_u), q_u(x_u))$  est indépendante de l'arbre de dépendances à bi-probabilité peau et non peau (Meilleur Arbre Recouvrant) considéré. En outre, la divergence de Kullback-Leibler est non-négative. Il suit que la maximisation de la mesure de la divergence de Kullback-Leibler  $KL(p(x|T), q(x|T))$  (le critère d'optimisation du Meilleur Arbre Recouvrant) est équivalente à la maximisation du poids de l'arbre  $T$ , exprimé par la formule ci-après :

$$\sum_{(u \sim v) \in T} \{KL(p_{uv}(x_u, x_v), q_{uv}(x_u, x_v)) - KL(p_u(x_u), q_u(x_u)) - KL(p_v(x_v), q_v(x_v))\} \quad (3.17)$$

À l'issue de ces calculs, nous pouvons présenter le théorème (2) qui caractérise les probabilités modélisées par le Meilleur arbre Recouvrant et approximant les distributions de probabilités réelles peau et non peau :

**Théorème 2.** .

*Les probabilités modélisées par le Meilleur Arbre Recouvrant  $T$ ,  $p(x|T)$  et  $q(x|T)$ , sont respectivement les approximations optimums des distributions de probabilités réelles peau et non peau,  $p(x)$  et  $q(x)$ , au sens du critère donné par l'équation (3.10), si et seulement si le poids de l'arbre  $T$ , défini par :*

$$W^T = \sum_{(u \sim v) \in T} \{KL(p_{uv}(x_u, x_v), q_{uv}(x_u, x_v)) - KL(p_u(x_u), q_u(x_u)) - KL(p_v(x_v), q_v(x_v))\} \quad (3.18)$$

*est maximal.*

*Avec  $p_u$  et  $q_u$  sont respectivement les lois marginales de  $p$  et  $q$  pour un noeud donné, alors que  $p_{uv}$  et  $q_{uv}$  sont les lois marginales d'un couple de noeuds  $(x_u, x_v)$ .  $W^T$  désigne le poids de l'arbre  $T$ .*

Ce théorème résume la recherche de l'arbre de dépendances à bi-probabilité peau et non peau au sens du critère d'optimisation  $\zeta$  (exprimé par l'équation (3.10)), aussi dit Meilleur Arbre Recouvrant, à la recherche de l'arbre de recouvrement maximal dont le poids est donné par l'équation (3.18).

Il convient que pour chaque couple de noeuds  $u$  et  $v$  distincts du graphe  $G$  représentant le vecteur des observations  $x = (x_1, x_2, \dots, x_n)$ , l'arc reliant  $u$  et  $v$  soit pondéré par l'expression :

$$KL(p_{uv}(x_u, x_v), q_{uv}(x_u, x_v)) - KL(p_u(x_u), q_u(x_u)) - KL(p_v(x_v), q_v(x_v)) \quad \forall (u, v) \in A \quad (3.19)$$

Informellement, cette dernière expression mesure la similarité entre les distributions de probabilités marginales des probabilités approximées  $p(x|T)$  et  $q(x|T)$  au niveau de chaque noeud et de chaque couple de noeuds du graphe. Ainsi, la mesure de similarité entre deux distributions de probabilités sur un arbre de dépendances à bi-probabilité (comme exprimé par l'équation (3.9)) est équivalente à la somme des mesures de similarités et dissimilarités entre les probabilités marginales de ces distributions au niveau de tous les noeuds et entre deux noeuds voisins de cet arbre.

La Figure 3.5 illustre quelques entrées (images couleurs) et sorties (images aux niveaux de gris) deux détecteurs de la peau fondés sur deux types d'arbres de dépendances à bi-probabilité. Les sorties sur la deuxième et troisième colonne correspondent respectivement à celles du détecteur fondé sur l'arbre Simple et celui basé sur le Meilleur Arbre Recouvrant.

En premier, nous tenons à noter qu'aucun post traitement morphologique (érosion, délation, etc.) n'est appliqué sur les images représentées en sorties des deux détecteurs utilisés. Nous constatons que les résultats obtenus par le modèle du Meilleur Arbre Recouvrant présentent une amélioration remarquable dans la différenciation entre les pixels peau et non peau des images analysées, et ce vis-à-vis du modèle de l'arbre de dépendances à bi-probabilité peau et non peau Simple. Cela prouve que le critère d'optimisation considéré par ce modèle du Meilleur Arbre Recouvrant exploite au mieux les interactions entre les deux classes peau et non peau. Il reste encore quelques pixels, en particulier au niveau des contours des zones de la peau, qui sont mal classés par le modèle du Meilleur Arbre recouvrant.



FIGURE 3.5 – Des entrée (première colonne à gauche) et des sorties de deux détecteurs de peau fondés sur deux types d'arbres de dépendances à bi-probabilité. Les images aux niveaux de gris de la deuxième et de la troisième colonne correspondent respectivement à celles du détecteur fondé sur l'arbre Simple et celui basé sur le Meilleur Arbre Recouvrant.

## 3.5 Vers un modèle d'arbres à bi-probabilité peau et non peau Discriminatoire

### 3.5.1 Comparaison des deux modèles d'arbres à bi-probabilité peau et non peau proposés

Nous avons proposé le modèle d'arbre à bi-probabilité peau et non peau dans l'objectif de définir un modèle discriminatoire d'arbre indexant les probabilités peau et non peau. En particulier, au lieu de considérer chacune des deux classes étudiées l'une indépendamment de l'autre, ce modèle assure le traitement des interactions entre elles. Ceci est principalement assuré par le fait que l'arbre à bi-probabilité indexe leurs deux distributions de probabilités à la fois. D'ailleurs la structure de l'arbre proposé, est déterminée à partir d'un apprentissage sur deux bases dont chacune comporte des exemples variés de la classe qu'elle représente (classe peau ou classe non peau). Aussi, le critère d'optimisation qui permet de définir l'arbre de dépendances à bi-probabilité le plus adéquat, doit permettre de modéliser les distributions de probabilités qui estiment au mieux les probabilités peau

et non peau réelles.

Dans la quête de l'arbre de dépendances à bi-probabilité peau et non peau qui discrimine au mieux les pixels peau et non peau d'une image, nous avons envisagé :

- dans un premier temps, l'arbre de dépendances à bi-probabilité peau et non peau Simple ;
- et dans un deuxième temps, le Meilleur Arbre Recouvrant.

Concernant le premier modèle, le critère d'optimisation porte sur le calcul de la similarité entre la probabilité indexée et sa distribution de probabilité approximée par ce modèle, et cela aussi bien dans le cas de la classe peau que dans le cas de la classe non peau. Physiquement, cette mesure de similarité permet de déterminer l'intra classe entre les distributions réelles et celle approximées. Quant à l'interclasse entre les distributions de probabilités modélisées par l'arbre en question, elle reste assurée par le fait que l'arbre modélise les distributions peau et non peau à la fois. Ainsi, pour que les probabilités modélisées par l'arbre de dépendances à bi-probabilité Simple soient optimales, il faut que la somme des mesures de similarités (cité précédemment) soit minimale. C'est exactement ce qu'est traduit par le critère d'optimisation du premier modèle proposé.

Quant au deuxième modèle, l'accent est porté sur la mesure de similarité, en particulier celle mesurée entre les distributions de probabilités modélisées par l'arbre. Contrairement au modèle dit arbre de dépendances à bi-probabilité Simple, le Meilleur arbre recouvrant répond au problème d'intra classe entre peau et non peau par le biais de l'utilisation d'un arbre unique. Le critère d'optimisation de cet arbre permet de déterminer les probabilités peau et non peau optimales par la maximisation de la distance entre les distributions approximées. Le choix d'un tel critère est justifié par le fait que la problématique de la détection de la peau dans l'image couleur est due au problème d'interclasse, comme il l'a été souligné dans l'introduction de ce mémoire.

Nous constatons aussi bien dans le premier modèle que dans le deuxième, que les critères d'optimisations utilisés sont limités. Ils doivent être étendus afin que l'arbre de dépendances à bi-probabilité peau et non peau modélise au mieux les probabilités qu'il indexe. En fait, les limitations des deux modèles proposés résident dans le fait que : la gestion du problème d'interclasse entre peau et non peau n'est pas explicitée par le critère d'optimisation dans le premier modèle, alors que dans le deuxième modèle c'est plutôt la gestion du problème d'intra classe entre ces deux classes qui n'est pas explicitée. Une formulation mathématique

couplant les contraintes sur ces deux problèmes, garantira de prédire les performances des distributions de probabilités estimées.

### 3.5.2 Modèle d'arbre de dépendances à bi-probabilité peau et non peau Discriminatoire

Suite aux constats reportés dans le paragraphe précédent, nous proposons un critère d'optimisation qui permet de gérer explicitement aussi bien l'interclasse entre les distributions de probabilités peau et non peau modélisées que l'intra classe entre chaque probabilité indexée que celle modélisée au niveau de chaque classe (peau ou non peau). Il paraît d'ores et déjà convenable de définir l'arbre de dépendances à bi-probabilité peau et non peau comme suit :

**Définition 9.** *Arbre de dépendance à bi-probabilité peau et non peau Discriminatoire*

*C'est l'arbre de dépendances indexant à la fois la probabilité peau et non peau tel que :*

1. *étant donnée une classe (peau ou non peau), la similarité entre la distribution indexée et celle modélisée est maximale ;*
2. *la similarité entre les deux distributions de probabilités modélisées est minimale.*

Pour comprendre la nature du critère d'optimisation adéquat pour déterminer l'arbre de dépendances à bi-probabilité peau et non peau Discriminatoire, nous présentons l'exemple suivant.

Exemple : Considérons deux ensembles  $A$  et  $B$  représentés par la Figure 3.6.

La probabilité jointe des deux ensembles  $A$  et  $B$ , est égale à :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (3.20)$$

Par analogie à cet exemple, le critère d'optimisation doit considérer de part et d'autre l'inter et l'intra classe entre les classes peau et non peau. Concernant les mesures des similarités, nous considérons toujours la métrique informationnelle divergence de Kullback Leibler. Nous supposons que les ensembles  $A$  et  $B$  représentent respectivement les classes peau et non peau. Ainsi, nous pouvons écrire formellement :

$$P(A) \Leftrightarrow KL(p(x), p(x|T)) \quad (3.21)$$

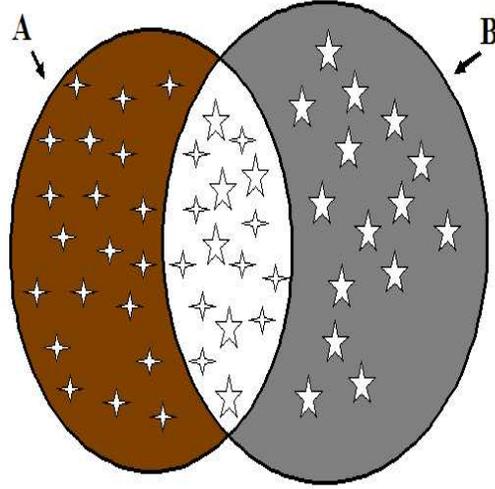


FIGURE 3.6 – Représentation graphique de deux ensembles  $A$  et  $B$ .

$$P(B) \Leftrightarrow KL(q(x), q(x|T)) \quad (3.22)$$

$$P(A \cap B) \Leftrightarrow KL(p(x|T), q(x|T)) \quad (3.23)$$

Il suit :

$$P(A \cup B) \Leftrightarrow KL(p(x), p(x|T)) + KL(q(x), q(x|T)) - KL(p(x|T), q(x|T)) \quad (3.24)$$

Lorsque les termes à droite des équations (3.21) et (3.22) sont minimaux, ils traduisent formellement le point (1) de la définition 9. Quant au point (2) de la même définition, il est exprimé par la maximisation du terme à droite de l'équation (3.23).

En Partant du principe que la divergence de Kullback Leibler soit non-négative, nous déduisons que la fonction  $\phi_3$  caractéristique du critère d'optimisation peut être formulée par le terme à droite de l'équation (3.24). D'où :

$$\phi_3(T) = KL(p(x), p(x|T)) + KL(q(x), q(x|T)) - KL(p(x|T), q(x|T)) \quad \forall T \subset G \quad (3.25)$$

A ce stade, nous pouvons définir le théorème caractéristique de l'arbre de dépendances à bi-probabilité peau et non peau, donné par le théorème (3).

**Théorème 3.** .

Un arbre de dépendances à bi-probabilité peau et non peau  $\tilde{T}$  est discriminatoire si et seulement si la contrainte ci-après est vérifiée :

$$\tilde{T} = \arg \min_{T \subset G} \{KL(p(x), p(x|T)) + KL(q(x), q(x|T)) - KL(p(x|T), q(x|T))\} \quad (3.26)$$

Dans cette section nous avons défini le critère d'optimisation qui peut décrire un modèle d'arbre discriminatoire indexant les probabilités peau et non peau. Dans la section ci-dessous nous décrivons la manière dont un tel arbre sera sélectionné à partir d'un graphe de dépendances à bi-probabilité peau et non peau.

### 3.5.3 Sélection de l'arbre de dépendances à bi-probabilité peau et non peau Discriminatoire

Considérons  $G(N, A)$  un graphe de dépendances à bi-probabilité peau et non peau, l'arbre discriminatoire des distributions peau et non peau, est l'arbre recouvrant de ce graphe qui vérifie la contrainte exprimée par le théorème (3). Pour construire le graphe en question, il est indispensable de déterminer les pondérations de ses arcs. A cette fin, nous devons exploiter la formule donnée par l'équation (3.26).

L'équation (3.26) offre l'arbre optimal qui maximise la fonction caractéristique du modèle proposé ( $\phi_3$ ). Remarquons que cette dernière n'est qu'une simple combinaison linéaire des deux fonctions caractéristiques ( $\phi_1$  et  $\phi_2$ ) des critères d'optimisation des premiers deux modèles proposés. Ainsi, les pondérations du graphe  $G$  peuvent être déduites par l'exploitation de :

$$\phi_3(T) = \phi_1(T) - \phi_2(T) \quad \forall T \subset G \quad (3.27)$$

Suite aux calculs présentés précédemment, et grâce aux équations (3.6) et (3.28), nous concluons que la formule donnée à (3.27) s'écrit aussi sous la forme :

$$KL(p(x), p(x|T)) + KL(q(x), q(x|T)) - KL(p(x|T), q(x|T)) = - \overbrace{\sum_{x_u \in N} KL(p_u(x_u), q_u(x_u))}^A$$

$$+ \overbrace{\sum_{x \in N} q(x) \log q(x) - \sum_{u \in N} q(x_u) \log q(x_u) + \sum_{x \in N} p(x) \log p(x) - \sum_{u \in N} p(x_u) \log p(x_u)}^B$$

$$\begin{aligned}
 & - \overbrace{\sum_{(u \sim v) \in T} \{ KL(p_{uv}(x_u, x_v), p_u(x_u)p_v(x_v)) + KL(q_{uv}(x_u, x_v), q_u(x_u)q_v(x_v)) \}}^C + \\
 & \underbrace{\sum_{(u \sim v) \in T} \{ KL(p_u(x_u), q_u(x_u)) + KL(p_v(x_v), q_v(x_v)) - KL(p_{uv}(x_u, x_v), q_{uv}(x_u, x_v)) \}}_D \quad (3.28)
 \end{aligned}$$

Etant donné que les deux termes  $A$  et  $B$  ne dépendent pas de l'arbre de dépendance à bi-probabilité  $T$ , la solution de l'arbre de dépendances à bi-probabilité Discriminatoire est déterminée par les deux termes  $C$  et  $D$ . Ces derniers expriment des sommes sur tous les couples des noeuds d'un arbre  $T$ . Il est donc clair que le graphe de dépendances à bi-probabilité peau et non peau est pondéré par les poids suivants :

$$\begin{aligned}
 & KL(p_{uv}(x_u, x_v), p_u(x_u)p_v(x_v)) + KL(q_{uv}(x_u, x_v), q_u(x_u)q_v(x_v)) - KL(p_u(x_u), q_u(x_u)) \\
 & - KL(p_v(x_v), q_v(x_v)) + KL(p_{uv}(x_u, x_v), q_{uv}(x_u, x_v)) \quad \forall (u \sim v) \in A \quad (3.29)
 \end{aligned}$$

D'où un deuxième théorème caractéristique de l'arbre discriminatoire recherché.

**Théorème 4.** .

*Un arbre de dépendances à bi-probabilité peau et non peau  $\tilde{T}$  est dit Discriminatoire pour les deux classes peau et non peau si et seulement si son poids donné par*

$$\begin{aligned}
 W^T = & \sum_{(u \sim v) \in T} \{ KL(p_{uv}(x_u, x_v), p_u(x_u)p_v(x_v)) + KL(q_{uv}(x_u, x_v), q_u(x_u)q_v(x_v)) + \\
 & KL(p_{uv}(x_u, x_v), q_{uv}(x_u, x_v)) - KL(p_u(x_u), q_u(x_u)) - KL(p_v(x_v), q_v(x_v)) \} \quad (3.30)
 \end{aligned}$$

*est maximal.*

Ce dernier théorème explicite comment l'inter et l'intra classe entre les distributions de peau et non peau modélisées par l'arbre de dépendances à bi-probabilité peau et non peau Discriminatoire, doivent être gérés aussi bien au niveau de chaque noeud de cet arbre que au niveau de chaque couple de noeuds voisins dans celui-ci.

La Figure 3.7 illustre quelques entrées (images couleurs) et sorties (images aux niveaux de gris) de deux détecteurs fondés sur deux types d'arbres de dépendances à bi-probabilité.

Les images aux niveaux de gris représentées sur la deuxième et troisième colonne correspondent respectivement aux résultats obtenus par le détecteur de peau fondé sur le Meilleur Arbre Recouvrant et celui basé sur l'arbre Discriminatoire.



FIGURE 3.7 – Des entrée (première colonne à gauche) et des sorties de deux détecteurs de peau fondés sur deux types d'arbres de dépendances à bi-probabilité. Les sorties sur la deuxième et troisième colonne correspondent respectivement à celles du détecteur fondé sur le Meilleur Arbre Recouvrant et celui basé sur l'arbre Discriminatoire.

Les zones peau et non peau sont d'avantage séparées dans les sorties du modèle de l'arbre de dépendance à bi-probabilité peau et non peau Discriminatoire, et cela vis-à-vis des sorties données par le modèle du Meilleur Arbre Recouvrant. La différence entre les résultats obtenus par les deux modèles est principalement constatée au niveau des contours des zones peau.

Concernant les arbres de dépendances à bi-probabilité peau et non peau définis dans le présent chapitre, nous constatons que partant du modèle de l'arbre Simple passant par le Meilleur Arbre Recouvrant et arrivant à l'arbre Discriminatoire, les résultats des images présentées montrent que les pixels sont d'avantage mieux classés. Cela affirme que sur ces images, les frontières entre les deux classes peau et non peau sont mieux tracées par le dernier modèle proposé.

Une évaluation expérimentale, en termes de qualité et de quantité des résultats obtenus par les trois modèles proposés, sera présentée au chapitre 5.

## 3.6 Arbres de dépendances à bi-probabilité peau et non peau

### 3.6.1 Procédures

Dans les section précédentes, nous avons proposé trois critères d'optimisation afin de caractériser les distributions peau et non peau. Chacun de ces critères a donné naissance à un modèle de détection de la peau indexant conjointement la probabilité peau et la probabilité non peau. La procédure ci-après récapitule les démarches nécessaires pour la détermination d'un arbre de dépendances à bi-probabilité peau et non peau. La base des exemples de la classe peau et celle de non peau est référencée par l'ensemble :

$$\{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$$

avec  $x^i$  désigne l'image  $i$  et  $y^i$  son masque associé.

**Procédure 1.** *Sélection de l'arbre de dépendances à bi-probabilité peau et non peau.*

– Entrée : Donnée  $\{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$ .

– Etapas :

1. Définir le système de voisinage à considérer selon la formule (3.1).
2. Construire le vecteur d'observation  $x$  issu de chaque bloc de voisinage.
3. Construire  $G(\mathcal{N}, \mathcal{A})$  le graphe non orienté correspondant à  $x$  ; avec
  - $\mathcal{N}$  : l'ensemble des noeuds. Chaque élément de  $x$  est un noeud.
  - $\mathcal{A}$  : l'ensemble des arcs déduit par le système de voisinage utilisé.
4. Pour chaque couple de noeuds  $u$  et  $v$  du graphe, utiliser un estimateur classique pour calculer les probabilités empiriques :  $p_u$  et  $p_{uv}$  (respectivement  $q_u$  et  $q_{uv}$ ) de  $p$  (respectivement  $q$ ) :

$$p_u(x_u = i) = f_i^1(x_u) \quad (3.31)$$

$$q_u(x_u = i) = f_i^2(x_u) \quad (3.32)$$

$$p_{uv}(x_u = i, x_v = j) = f_{ij}^1(x_u, x_v) \quad (3.33)$$

$$q_{uv}(x_u = i, x_v = j) = f_{ij}^2(x_u, x_v) \quad (3.34)$$

où pour  $m = 1, 2$ ,  $f_{ij}^m(x_u, x_v)$  est la fréquence jointe de  $x_u = i$  et  $x_v = j$  étiquetées  $m$ . Et  $f_i^m(x_u)$  est la fréquence de  $x_u = i$  étiquetée  $m$ .

5. Calculer le coût des arcs défini par le type d'arbre de dépendances à déterminer.
6. Utiliser l'algorithme de Kruskal pour déterminer un arbre de recouvrement optimal.

– Sortie : L'arbre de dépendances à bi-probabilité peau et non peau.

Les différents arbres de dépendances à bi-probabilité peau et non peau proposés (Simple, Meilleur Arbre Recouvrant et Discriminatoire) peuvent être déterminés à partir de la procédure ci-dessus. La différence entre leurs procédures se situe au niveau de l'étape (5).

En fait, selon l'arbre à déterminer le coût d'un arc de cet arbre est égal à :

– Arbre Simple :

$$KL(p_{uv}(x_u, x_v), p_u(x_u)p_v(x_v)) + KL(q_{uv}(x_u, x_v), q_u(x_u)q_v(x_v)) \quad (3.35)$$

– Meilleur Arbre Recouvrant :

$$KL(p_{uv}(x_u, x_v), q_{uv}(x_u, x_v)) - KL(p_u(x_u), q_u(x_u)) - KL(p_v(x_v), q_v(x_v)) \quad (3.36)$$

– Arbre Discriminatoire :

$$KL(p_{uv}(x_u, x_v), p_u(x_u)p_v(x_v)) + KL(q_{uv}(x_u, x_v), q_u(x_u)q_v(x_v)) + KL(p_{uv}(x_u, x_v), q_{uv}(x_u, x_v)) \\ - KL(p_u(x_u), q_u(x_u)) - KL(p_v(x_v), q_v(x_v)) \quad (3.37)$$

### 3.6.2 Inférences

Dans la section précédente, nous avons décrit la phase de la modélisation des distributions de probabilités peau et non peau, et cela pour les trois modèles proposés. Une fois la phase de la modélisation est achevée, il est nécessaire de calculer les inférences pour mesurer l'incertitude qu'un pixel soit de la classe peau ou de la classe non peau.

Considérons  $y_s$  l'état (peau ou non peau) du pixel  $s$  à analyser. Afin de calculer par chacun des trois modèles d'arbres présentés auparavant l'état du pixel  $s$  sachant le vecteur des observations  $x$ , nous faisons appel à la formule de Bayes. Ainsi nous obtenons :

$$P(y_s = j|x) = \frac{P(y_s = j)P(x|y_s = j)}{P(x)} \quad , \quad j = 1, 2. \quad (3.38)$$

Tel que nous avons :

$$P(x) = \sum_{y_s=1}^2 P(x, y_s) = \sum_{i=1}^2 P(x|y_s = i)P(y_s = i)$$

où

$$P(x|y_s = 2) \approx q_T(x) = \prod_{(u \sim v) \in T} \frac{q_{uv}(x_u, x_v)}{q_u(x_u)q_v(x_v)} \prod_{x_u \in V} q_u(x_u) \quad (3.39)$$

$$P(x|y_s = 1) \approx p_T(x) = \prod_{(u \sim v) \in T} \frac{p_{uv}(x_u, x_v)}{p_u(x_u)p_v(x_v)} \prod_{x_u \in V} p_u(x_u) \quad (3.40)$$

$$P(y_s = 2) \approx q_T(y_s = 2) \quad (3.41)$$

$$P(y_s = 1) \approx p_T(y_s = 1) \quad (3.42)$$

Les éléments des équations (3.39), (3.40), (3.41), et (3.42) sont déjà calculés à l'étape (4) de la procédure de sélection de l'arbre de dépendances à bi-probabilité peau et non peau, et cela pour chacun des trois modèles présentés dans ce chapitre.

### 3.7 Discussion

Les modèles d'arbres indexant des probabilités sont classés en deux catégories : des arbres dont la structure est fournie a priori et ceux dont la structure est définie via un ensemble de données. Nous privilégions la deuxième catégorie d'arbres du fait qu'elle reflète des connaissances sur le domaine étudié. Or, dans le cadre de la classification à titre d'exemple, les modèles d'arbres proposés dans ce contexte, modélisent une distribution de probabilité en se restreignant aux exemples de la classe représentée. Nous nous limitant à la modélisation des distributions de probabilités peau et non peau.

La détection de la peau dans les images couleurs est une problématique ardente qui fait l'objet de plusieurs études de recherches. Dans l'objectif de déterminer un modèle robuste de la peau, nous exploitons les modèles d'arbres probabilistes. Or, le problème d'interclasse est le noyau de la problématique de la détection de la peau. De ce constat, il paraît que l'apprentissage de la structure de l'arbre modélisant la distribution de la peau (respectivement non peau), uniquement à partir des exemples de la peau (respectivement non peau) n'est pas suffisant pour une meilleure modélisation. Il est plus adéquat que la

structure d'arbre modélisant la distribution de peau (respectivement non peau), prenne en considération le comportement de la distribution non peau (respectivement peau). Ainsi, dans un cadre de classification supervisée, nous avons proposé un modèle d'arbre indexant les deux probabilités peau et non peau, à la fois. Ce modèle, que nous appelons arbre de dépendances à bi-probabilité peau et non peau, exploite notre vision pour une structure d'arbre mettant l'accent sur les interactions distinguant les deux distributions modélisées. La structure de l'arbre de dépendance à bi-probabilité peau et non peau, est fournie à partir d'un ensemble d'exemples de peau et non peau étiquetés. Ainsi, malgré l'hétérogénéité des données, nous disposons des connaissances sur les exemples de chaque distribution. Cela garantit que les distributions peau et non peau modélisées sont conformes aux distributions réelles.

En outre, l'arbre recouvrant optimal a prouvé sa performance dans plusieurs domaines d'activités. Nous avons adapté ce principe dans la définition du modèle de l'arbre proposé. L'arbre de dépendances à bi-probabilité peau et non peau est l'arbre recouvrant d'un graphe représentant les données des deux distributions. Toutefois, la difficulté réside dans la détermination du critère d'optimisation convenable pour que l'arbre approxime au mieux les probabilités peau et non peau réelles. Une partie importante de ce chapitre a été consacrée à cette fin. Au fil de plusieurs propositions concernant le critère d'optimisation séparant au mieux les pixels peau et non peau, nous avons défini un ensemble de modèles d'arbres indexant deux probabilités peau et non peau. Les différentes propositions considèrent des problèmes d'interclasse et d'intra classe entre les distributions à modéliser. La gestion des deux derniers problèmes est explicitement ou implicitement assurée par les différents modèles. Suite aux différentes propositions du critère d'optimisation, l'arbre de dépendances à bi-probabilité peau et non peau Discriminatoire reste formellement le modèle robuste pour l'indexation des deux probabilités peau et non peau.

Bien qu'une gestion de l'interclasse et l'intra classe par une formulation mathématique soit la plus appropriée, nous pensons que les différents modèles d'arbres proposés dans ce chapitre, améliorent les modèles d'arbres existants pour la détection de la peau. Nous cédon la place aux expérimentations et aux études comparatives des différents modèles(chapitre 5) pour sélectionner l'arbre de peau le plus robuste.



---

## MODÈLES PROBABILISTES INDEXÉS PAR DES MÉLANGES OU DES COMBINAISONS DES ARBRES

Dans le chapitre précédent, nous avons défini trois différents modèles d'arbres fondés sur le principe de la co-modélisation des probabilités peau et non peau par un arbre unique. En outre, le concept d'arbre recouvrant optimal (maximal ou minimal) est invoqué dans la définition de chaque modèle. Cet arbre a trouvé son application dans plusieurs secteurs d'activités tel que les réseaux informatiques et les réseaux de transports. Cependant, l'un des problèmes majeurs auxquels, ce genre d'arbres se retrouve confronté, est la possible multiplicité des arbres recouvrants optimaux (ArbreROpts) sur un même graphe.

Dans une première section, le présent chapitre met l'accent sur le problème de multiplicité d'arbres recouvrants optimaux ainsi que sur les probabilités qu'ils modélisent. Les algorithmes classiques tels que Prim ou Kruskal, ne permettent nullement de différencier entre les différents ArbreROpts possibles. En effet, sur un graphe, ils sélectionnent un ArbreROpt aléatoire. De plus, le choix de cet arbre peut varier d'un algorithme à un autre. Notre travail consiste à déterminer l'ArbreROpt idéal modélisant les distributions de probabilités peau et non peau. Pour cela, la deuxième section présente une méthode fondée sur le mélange conjoint, et des ArbreROpts et de leurs probabilités modélisées. Des études théoriques sur le mélange de quatre modèles d'arbres présentés à la section 4.2.1, prouvent les améliorations apportées par une telle méthode par rapport aux méthodes fondées sur l'utilisation d'un unique ArbreROpt. Toutefois, pour quelques types d'arbres particuliers, il s'est avéré nécessaire de procéder à une autre alternative que le mélange. Ainsi, une deuxième solution est proposée dans la quatrième section du présent chapitre.

## 4.1 Ensemble des arbres recouvrants optimaux

### 4.1.1 Multiplicité des arbres de recouvrement optimaux d'un graphe

Un graphe de dépendances représentant des données et modélisant un aspect spécifique de celles-ci, permet de déterminer les différentes corrélations et interactions entre les données. Par ailleurs, un arbre recouvrant optimal (maximal ou minimal) de ce graphe, définit une structure de dépendance et une distribution particulières des données. Les ArbreROpts ont trouvé leurs utilisations dans plusieurs domaines de recherches sous l'onglet de l'optimisation. Néanmoins, un ArbreROpt n'est pas toujours unique sur un graphe donné. En effet, dans le cas le plus simple d'un graphe dont les pondérations des arcs sont identiques, tout arbre recouvrant est un ArbreROpt ; ou encore comme illustre l'exemple donné par la Figure 4.1, trois arbres optimaux sont tirés du graphe représenté.

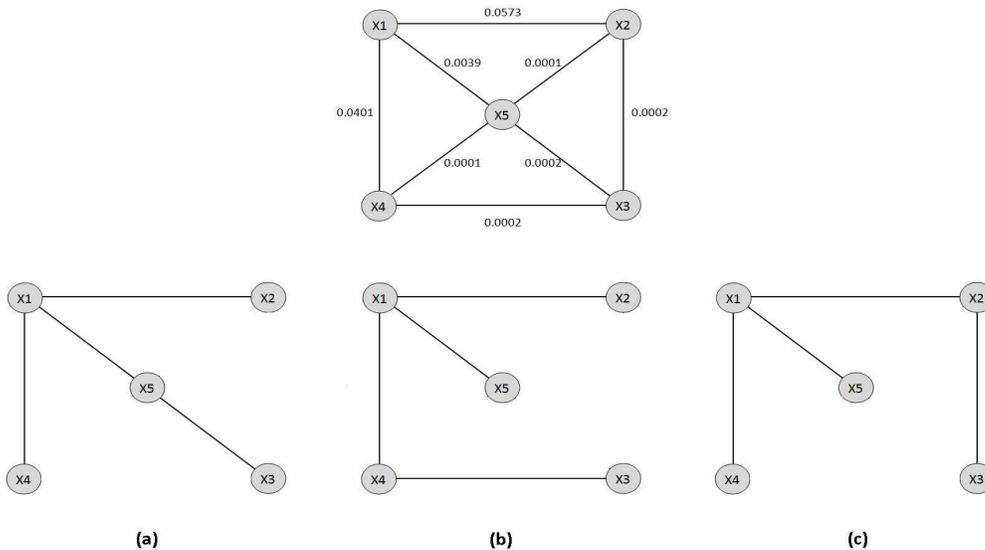


FIGURE 4.1 – Un graphe de dépendances (en haut) et ces arbres de recouvrements optimaux (en bas).

Considérons un vecteur des observations  $x = (x_1, x_2, x_3, x_4, x_5)$  dont la loi de probabilité est notée  $p$ ; les  $x_i$  ( $i=1, 2, \dots, 5$ ) sont des variables aléatoires discrètes. Supposons que le graphe représentant  $x$  ainsi que les dépendances entre les variables discrètes, soit celui donné par la Figure 4.1. Supposons aussi, que l'ArbreROpt à sélectionner soit obtenu par

la maximisation d'un certain critère  $\zeta$ . Il parvient que trois arbres de recouvrement maximaux représentés aux Figures 4.1.a, 4.1.b et 4.1.c, soient possibles. Chacun de ces arbres détermine une structure de dépendances bien particulière et modélise une approximation de la distribution de probabilité  $p$  au sens du même critère  $\zeta$ . Les probabilités données par les différents ArbreROpts (a), (b) et (c) sont respectivement :

$$p^a(x) = p(x_1)p(x_2|x_1)p(x_4|x_1)p(x_5|x_1)p(x_3|x_5)$$

$$p^b(x) = p(x_1)p(x_2|x_1)p(x_4|x_1)p(x_5|x_1)p(x_3|x_4)$$

$$p^c(x) = p(x_1)p(x_2|x_1)p(x_4|x_1)p(x_5|x_1)p(x_3|x_2)$$

ces trois probabilités ne sont pas nécessairement égales même si elles approximent la même probabilité  $p$ . Cet exemple démontre que chaque arbre optimal modélise, selon sa structure, une approximation particulière d'une distribution de probabilité.

Dans un graphe indexant une probabilité ou plus, chaque arbre de recouvrement optimal répond à un critère de maximisation ou de minimisation bien spécifié. D'ailleurs, cet arbre est un arbre de dépendances qui permet de mettre en évidence les importantes interactions entre les données afin d'offrir la meilleure approximation de la distribution de celles-ci au sens du critère utilisé. Or dans la plupart des cas d'optimisations basés sur les ArbresROpts, ce qui est souvent demandé ce n'est pas un ArbreROpt mais plutôt l'arbre "idéal". L'arbre " idéal ", c'est l'ArbreROpt ayant la meilleure structure et qui approxime en mieux la distribution réelle des données. Néanmoins, les descriptions classiques des algorithmes de calcul des ArbresROpts comme Kruskal ou Prim, ne sont pas déterminants. Autrement dit, afin de déterminer l'arbre recouvrant minimal par exemple, ces algorithmes considèrent les arcs de poids minimaux dans l'ordre en vérifiant qu'aucun cycle n'est formé. Mais lorsqu'un choix entre au moins deux arcs ayant la même pondération se présente, aucune règle exigeant la sélection d'un arc plutôt qu'un autre n'est spécifiée. C'est pourquoi, un problème ardu, à savoir la multiplicité du choix de l'ArbresROpt, s'impose et deux questions apparaissent : (1) quel est l'arbre optimal à choisir et par conséquence la meilleure structure d'arbre d'approximation à considérer parmi tout les ArbresROpts possibles ? et (2) faut-il tenir compte de tous ces arbres recouvrants optimaux ou simplement se contenter de choisir un arbre aléatoirement, puisqu'ils répondent tous au même critère d'optimisation ?

Dans un premier temps nous nous intéressons à la détermination de tous les arbres de recouvrement optimaux d'un graphe avant de répondre dans la section 4.2 aux deux questions soulevées.

### 4.1.2 Extension des arbres de recouvrement optimaux

Nous soulignons deux points importants dans le cadre de la sélection des arbres de recouvrement optimaux :

- Les arbres de recouvrements optimaux sont déterminés au sens d'un critère ou des critères d'optimisation. Toutefois, la traduction formelle exacte de ces contraintes permettant de définir la sélection des arbres à définir n'est pas toujours garantie. En effet, les propriétés caractérisant l'arbre de dépendances optimal, dépendent de l'application en vue de réaliser et du secteur d'activité en question. Ces propriétés peuvent être parfois difficiles à quantifier ou encore impossible à connaître à priori, en particulier les plus discriminantes entre elles.
- Dans le cas des variables aléatoires discrètes et à cause de la discontinuité, les mesures calculées et par la suite le poids optimum de l'arbre retenu peut être très légèrement différent des mesures réelles.

Dans la perspective de pallier ces deux inconvénients, nous introduisons la notion de "l'erreur de tolérance" que nous définissons comme suit :

**Définition 10.** *Erreur de tolérance*

*C'est la marge d'erreur maximale que l'utilisateur estime être tolérable afin d'ajuster ses calculs.*

Considérons un graphe représentant un vecteur des observations. Par une recherche exhaustive, nous dressons une liste de tous les arbres recouvrants de ce graphe. Ensuite, nous ordonnons ces arbres dans l'ordre décroissant (respectivement, croissant) si l'optimisation est au sens de la maximisation (respectivement, la minimisation). Si la différence des poids des arbres de la tête de la liste est insignifiante, c'est à dire inférieure à l'erreur de tolérance que nous nous donnons, nous gardons un ensemble plus large des arbres recouvrants optimaux mais dont il est certain que la meilleure structure d'arbre adéquate à l'application envisagée existe.

Revenons aux modèles probabilistes indexés par les arbres, dans ce cas de figure l'erreur de tolérance permet d'étendre l'ensemble des `ArbreROpts` susceptibles de présenter les meilleures modélisations d'une probabilité. Aussi, elle permet d'exploiter les importantes dépendances entre les données et de capitaliser les informations utiles qui peuvent être ignorées, dans l'objectif de déterminer l'idéale modélisation des probabilités.

### 4.1.3 Sélection des arbres optimaux d'un graphe

Intuitivement, une recherche exhaustive dans un graphe permettra de lister tous les arbres recouvrants possibles de ce dernier et il suffit de calculer et d'ordonner les poids de ces différents arbres afin de déterminer l'ensemble des arbres recouvrants optimaux. Néanmoins, énumérer tous les arbres recouvrants d'un graphe est algorithmiquement difficile. En effet, le problème de dénombrement d'arbres dans un graphe n'est pas récent et a été étudié par A. Cayley qui a prouvé dans [98] que pour un graphe complet à  $n$  noeuds, le nombre d'arbre recouvrant est égal à  $n^{n-2}$ .

Dans l'objectif de sélectionner les arbres recouvrants optimaux, nous utilisons un algorithme de N. KATOH et al., décrit dans [99], dit algorithme de *K-Best-Spanning-Trees*; où  $K$  est un nombre entier. Il sert à déterminer les  $K$  meilleurs arbres recouvrants. Pour enlever toute ambiguïté, il serait convenable de noter que la notion de meilleur ici ne fait pas référence au type d'arbre que nous avons proposé dans la section 3.4.2 mais plutôt à la notion d'optimisation au sens d'un critère fixé par l'utilisateur. A cet algorithme nous ajoutons la contrainte exprimée par l'équation (4.1). Celle-ci désigne formellement l'ensemble des arbres recouvrants optimaux (maximaux ou minimaux).

$$W^{T_i} = \arg \max_T (W^T), \quad 1 \leq i \leq k \quad (4.1)$$

Avec :

- $W^T$  est le poids de l'arbre recouvrant  $T$  ;
- $W^{T_i}$  est le poids du  $i^{eme}$  optimal arbre recouvrant.

En ce qui concerne l'erreur de tolérance, elle peut être introduite dans les calculs en considérant un nombre  $\varepsilon$  réel strictement positif et négligeable devant 1 ( $0 < \varepsilon \ll 1$ ) et en exigeant que les arbres à sélectionner par l'algorithme utilisé vérifient l'inégalité suivante :

$$|W^{T_1} - W^{T_i}| \leq \varepsilon \quad (4.2)$$

avec  $W^{T_1}$  est le poids d'un arbre recouvrant optimal que nous définissons et fixons au début de l'algorithme.

Nous attirons l'attention sur l'équation (4.3) que nous considérons comme hypothèse admise du moment que  $\varepsilon \ll 1$ .

$$W^{T_1} = W^{T_i} \quad , \quad \forall i \in \{2, 3, \dots, K\} \quad (4.3)$$

Notons  $\Theta^k$  l'ensemble des arbres recouvrants optimaux résultant dont le cardinal ( $|\Theta^k|$ ) est égal à  $K$ . Ainsi, dans la suite nous parlons de  $K$  arbres recouvrants optimaux pour désigner le nombre d'arbres sélectionnés.

## 4.2 En quête de l'arbre recouvrant "idéal"

### 4.2.1 De la multiplicité des arbres recouvrants optimaux vers l'arbre recouvrant "idéal"

La principale difficulté dans les approches fondées sur l'arbre recouvrant optimal réside dans la possible multiplicité de ce genre d'arbre dans un graphe. Cependant, bien que tous les `ArbreROpts` répondent au même critère d'optimisation, l'arbre "idéal", parmi eux, le mieux approprié à une application donnée encapsule d'autres propriétés qui peuvent être difficiles à quantifier. Le choix aléatoire d'un arbre parmi les `ArbreROpts` possibles revient à privilégier sans fondement une structure de dépendances entre les données d'un graphe au dépend des autres arbres. Dans un graphe donné, au vu de l'importance des différentes informations apportées par chacun des arbres recouvrants optimaux, une exploitation de ces connaissances permettra d'améliorer les résultats apportés par une approche ne considérant que l'un de ces arbres indépendamment des autres. Il peut donc être intéressant de s'orienter vers des approches combinant les différents `ArbreROpts` possibles. Une façon de prendre en compte simultanément la structure de dépendances et l'information probabiliste sur les différents `ArbreROpts`, est le modèle de mélange. Il apparaît en effet adapté d'assembler les connaissances issues de  $K$  distributions homogènes des arbres recouvrants optimaux dans l'objectif d'aboutir à l'arbre "idéal" approximant le modèle probabiliste le plus approprié.

Dans un cadre général, le modèle de mélange de probabilités a prouvé son efficacité dans l'estimation d'une probabilité, et cela en comparaison avec un modèle ne considérant qu'une probabilité singulière. En particulier, dans le cas de la détection de la peau, nous citons l'exemple du mélange des gaussiennes qui présente une supériorité par rapport au modèle d'une simple gaussienne. Cela fut la raison derrière l'idée d'un mélange des probabilités estimées par les arbres recouvrants optimaux. D'autant plus que les modèles indexés par les arbres ont présentés des résultats prometteurs dans l'approximation des lois de probabilités réelles. C'est pourquoi nous choisissons que le mélange des probabilités des arbres recouvrants optimaux soit modélisé par un arbre que nous appelons arbre du mélange. Nous pensons qu'au lieu de considérer un mélange classique quelconque de distributions de probabilités, il est plus adéquat de procéder à un mélange modélisé par un arbre de mélange. C'est arbre estime la meilleure probabilité du mélange des probabilités possibles, c'est ainsi nous que le considérons comme étant l'arbre "idéal". En effet, dans la partie de la théorie mathématique présentée ci-après, nous montrons que le choix d'un procédé à base d'un mélange d'arbres de recouvrement ou encore de dépendances optimaux permet d'approximer au mieux les probabilités réelles. En fait, nous pouvons considérer les autres probabilités des mélanges possibles comme des probabilités modélisées par des mélanges autres que ceux des `ArbreROpts`.

Considérons  $x = (x_1, x_2, \dots, x_n)$  un vecteur des observations de  $n$  variables discrètes dont le graphe représentant est  $G(N, A)$ . En procédons à un mélange conjoint, et des structures de dépendances des différents `ArbreROpts` et des probabilités approximées par ces arbres, la distribution de probabilité du mélange est exprimée comme suit :

$$P(x|T_{mix}) = \sum_{T \in \Theta(G)^k} \lambda(T)P(x|T) \quad (4.4)$$

où

- $T_{mix}$  : l'arbre du mélange, dit idéal ;
- $P(x|T_{mix})$  : la probabilité du mélange approximée par l'arbre du mélange  $T_{mix}$  ;
- $\lambda(T)$  : le coefficient du mélange de la probabilité associée à l'arbre  $T$ . Avec  $\lambda(T) \geq 0$  et  $\sum_{T \in \Theta(G)^k} \lambda(T) = 1$  ;
- $P(x|T)$  : la probabilité modélisée par l'arbre  $T$  ;
- $\Theta(G)^k$  : l'ensemble de tous les arbres recouvrants optimaux du graphe  $G$ , dont le cardinal est égal à  $K$ .

Pour apporter plus de rigueur, nous proposons une étude théorique justifiant les apports du mélange des arbres recouvrants optimaux par rapport à un seul arbre de la même famille. Or, concernant cette famille d'arbres, nous avons vu pour le moment deux types de distributions : (1) les arbres modélisant une distribution de probabilité que nous nommons arbre à uni-probabilité et (2) ceux modélisant deux distributions de probabilités à la fois dits arbres à bi-probabilité. Dans les deux sections qui succèdent, nous discutons le mélange de chacun de ces deux cas de figures en particulier pour l'indexation des probabilités peau et non peau par des :

- arbres de Chow Liu (définis à la section 2.4.2) ;
- arbre de dépendances à bi-probabilité peau et non peau Simple (défini à la section 3.3) ;
- Meilleur Arbre Recouvrant (présenté à la section 3.4) ;
- arbre de dépendances à bi-probabilité peau et non peau Discriminatoire (donné par la section 3.5).

## 4.2.2 Etude théorique du mélange des arbres de Chow Liu

Au début, précisons que les notations de la section précédente seront conservées. De plus, considérons :

- $P(x)$  : la distribution de probabilité réelle de  $x$  ;
- $S(G)$  : l'ensemble de tous les arbres recouvrants du graphe  $G$  ;
- $S^K(G)$  : un sous ensemble de  $S(G)$  dont le cardinal est égal à  $K$  ;
- $\Theta(G)^k$  : l'ensemble de  $K$  arbres recouvrants optimaux ;
- $W^T$  : le poids de l'arbre  $T$ .

Rappelons le critère d'optimisation sur lequel est fondé l'arbre optimal  $\tilde{T}$  de Chow Liu :

$$\tilde{T} = \arg \min_T KL(p(x), p(x|T)) \quad (4.5)$$

Notons que le graphe  $G$  est pondéré selon ce critère. Aussi, au sens du critère de Chow Liu, nommons tout arbre recouvrant de  $G$  qui ne satisfait pas l'équation (4.5) un "mauvais arbre". En particulier, nous définissons le "plus mauvais arbre" du graphe  $G$  comme suit :

**Définition 11.** *"Plus mauvais arbre" de  $G$*

*C'est l'arbre recouvrant  $T'$  de  $G$ , dont la divergence de KL entre la distribution de proba-*

bilité réelle et celle approximée par cet arbre, est maximal. Formellement :

$$T' = \arg \min_{T \in S(G)} \left( \sum_{i=1}^n \log P(x_i|T) \right) \quad (4.6)$$

Ainsi nous déduisons l'inégalité qui suit :

$$W^{T'} \leq W^T \quad , \quad \forall T \in S(G) \quad (4.7)$$

Bien encore, nous présentons le théorème suivant :

**Théorème 5.** .

*Une distribution de probabilité est mieux modélisée par l'arbre  $T_{mix}$ , obtenu par le mélange des arbres recouvrants de l'ensemble  $S^K$ , que par le "plus mauvais arbre" de ce mélange. Autrement écrit :*

$$KL(P(x), P(x|T_{mix})) \leq KL(P(x), P(x|T')) \quad , \quad T' \in S^K \quad (4.8)$$

La preuve de ce théorème est envoyée à l'annexe (A) section (A.1).

Pour le moment, il est évident que le mélange des arbres recouvrants du graphe  $G$  modélise en mieux la probabilité réelle, que la probabilité modélisée par le "plus mauvais" de ces arbres. De plus, lorsque l'ensemble des arbres recouvrants choisi est celui des ArbreROpts, le modèle du mélange de ces arbres est plus performant que celui du "plus mauvais" des arbres considérés qui est simplement un arbre recouvrant optimal de  $G$ . Nous concluons le corollaire suivant :

**Corollaire 1.** .

*La distribution de probabilité est mieux modélisée par l'arbre  $T_{mix}$  obtenu par le mélange de  $K$  ArbreROpts, que par n'importe quel ArbreROpt  $T^{op}$ . Cela est équivalent à :*

$$KL(P(x), P(x|T_{mix})) \leq KL(P(x), P(x|T^{op})) \quad (4.9)$$

A ce stade, il est certain que le mélange des ArbreROpts de Chow Liu soit théoriquement plus performant mais la question qui s'impose est la suivante : en procédant par le mélange d'arbre recouvrant, n'existe-t-il pas un autre ensemble d'arbres recouvrants de  $G$ , dont le mélange est plus performant que celui des ArbreROpts?

Afin de répondre à cette question, nous considérons, en plus de l'ensemble  $\Theta(G)^k$ , un ensemble  $S^K$  d'arbres recouvrants quelconque de  $G$  tel que au moins l'un de ses arbres n'est pas un arbre recouvrant optimal. Et nous définissons les lois des mélanges issues des deux ensembles d'arbres comme suit :

$$P(x|T_{mix}^{\Theta^K}) = \sum_{T \in \Theta^K} \lambda(T)P(x|T) \quad , \quad \sum_{T \in \Theta^K} \lambda(T) = 1 \quad (4.10)$$

$$P(x|T_{mix}^{S^K}) = \sum_{T \in S^K} \lambda'(T)P(x|T) \quad , \quad \sum_{T \in S^K} \lambda'(T) = 1 \quad (4.11)$$

où les coefficients  $\lambda(T)$  et  $\lambda'(T)$  minimisent respectivement les deux divergences de Kullback - Leibeler  $KL(P(x), P(x|T_{mix}^{\Theta^K}))$  et  $KL(P(x), P(x|T_{mix}^{S^K}))$ .

Evaluer les performances des deux distributions des probabilités des mélanges, revient à comparer les divergences de Kullback - Leibeler  $KL(P(x), P(x|T_{mix}^{\Theta^K}))$  et  $KL(P(x), P(x|T_{mix}^{S^K}))$ . A la suite d'un calcul développé (présenté à l'annexe (A) section (A.2) de ces deux distances, nous obtenons :

$$KL(P(x), P(x|T_{mix}^{\Theta^K})) \leq KL(P(x), P(x|T_{mix}^{S^K})) \quad (4.12)$$

De ce fait, le théorème suivant est déduit :

**Théorème 6.** .

*La meilleure probabilité modélisée par un mélange d'arbres recouvrants du graphe  $G$ , est obtenue par l'ensemble de  $K$  arbres recouvrants optimaux de celui-ci.*

Dans la présente section, nous avons mis au clair deux points cruciaux, à savoir :

- Pour  $S^K(G)$  l'ensemble de  $K$  arbres recouvrants quelconques du graphe  $G$ , le mélange des arbres de  $S^K(G)$  tel qu'il est proposé dans (4.4) doit présenter des résultats plus performants que ceux du "plus mauvais arbre" de cet ensemble ;
- Le mélange de  $K$  arbres recouvrants optimaux d'un graphe est plus performant que le mélange de  $K$  arbres recouvrants quelconque du même graphe.

Ces deux points prouvent que le modèle du mélange des  $K$  arbres recouvrants optimaux de Chow Liu que nous proposons est une nouvelle perspective qui s'ouvre pour améliorer les modèles probabilistes indexés par des arbres.

### 4.2.3 Etudes théoriques du mélange de trois types d'arbres à bi-probabilité peau et non peau

Les notations mathématiques de la section précédentes seront réutilisées ici dans l'objectif de montrer l'apport des mélanges de trois modèles d'arbres bien particuliers de l'ensemble des arbres à bi-probabilité peau et non peau, à savoir l'arbre Simple, le Meilleur Arbre Recouvrant et l'arbre Discriminatoire.

Remarquons que les pondérations du graphe  $G$  dépendra du modèle d'arbres considéré. Pour deux lois de probabilités peau et non peau  $p$  et  $q$ , leurs deux distributions modélisées par un arbre  $T$  seront notées  $p(\cdot|T)$  et  $q(\cdot|T)$ .

Les distributions de probabilités  $p(x|T_{mix}^{\Theta^K})$  et  $q(x|T_{mix}^{\Theta^K})$  modélisées par un arbre à bi-probabilité "idéal"  $T_{mix}$ , obtenu par le mélange de  $K$  ArbreROpts, seront notées :

$$p(x|T_{mix}^{\Theta^K}) = \sum \lambda(T)p(x|T) \quad (4.13)$$

$$q(x|T_{mix}^{\Theta^K}) = \sum_{T \in \Theta^K} \lambda(T)q(x|T) \quad (4.14)$$

$$\sum_{T \in \Theta^K} \lambda(T) = 1 \quad (4.15)$$

Il serait convenable de noter que les deux distributions de probabilités  $p(x|T_{mix}^{\Theta^K})$  et  $q(x|T_{mix}^{\Theta^K})$  possèdent des coefficients des mélanges identiques. Cela est expliqué par le fait que nous procédons à un mélange d'arbres à bi-probabilité et des distributions modélisées.

Il convient donc que pour chaque couple de distributions modélisées par un arbre, les mêmes pondérations sont affectées aux composantes du couple.

#### 4.2.3.1 Mélange des arbres de dépendances à bi-probabilité peau et non peau Simples

Soit  $\tilde{T}$  un arbre de dépendances à bi-probabilité peau et non peau Simple indexant les distributions de probabilités réelles notées respectivement  $p$  et  $q$ . Comme il a été présenté au chapitre précédent, l'arbre à bi-probabilité  $T$  est caractérisé par la minimisation de la somme des divergences de Kullback-Leibler définies d'un côté entre la distribution de probabilité  $p$  et sa probabilité approximée par l'arbre en question  $p(\cdot|\tilde{T})$  et d'un autre côté entre la probabilité réelle  $q$  et sa probabilité approximée  $q(\cdot|\tilde{T})$  par le même arbre. Formellement, la fonction caractéristique du critère d'optimisation de  $T$  est :

$$KL(p(x), p(x|T)) + KL(q(x), q(x|T)) \quad \forall T \subset G \quad (4.16)$$

La solution obtenue par la minimisation de cette dernière formule, détermine l'arbre  $\tilde{T}$ . Afin que le modèle du mélange des différents arbres à bi-probabilité peau et non peau Simples possibles, soit plus performant que celui où un seul arbre de la même sélection des arbres considérés est utilisé, il faut que l'inégalité qui suit soit vérifiée :

$$KL(p(x), p(x|T_{mix})) + KL(q(x), q(x|T_{mix})) \leq KL(p(x), p(x|\tilde{T})) + KL(q(x), q(x|\tilde{T})) \quad (4.17)$$

Or, suite à l'étude théorique menée à la section 4.2.2, nous déduisons d'après l'inégalité présentée en (4.9) que la contrainte (4.17) est valide. De plus par le même procédé que celui adopté dans la section 4.2.2, nous pouvons montrer que le meilleur mélange d'arbre à considérer dans le cas du graphe à bi-probabilité peau et non peau Simple, est celui des arbres de dépendances à bi-probabilité peau et non peau Simples possibles.

Pour conclure cette section, nous pouvons annoncer que le mélange et des arbres de dépendances à bi-probabilité peau et non peau Simple et des probabilités approximées par ces arbres, est une solution de l'arbre "idéal" qui indexe les probabilités peau et non peau.

#### 4.2.3.2 Mélange des Meilleurs Arbres Recouvrants

D'après la définition d'un Meilleur Arbre Recouvrant, ce dernier est fondé sur le critère suivant :

$$T^{op} = \arg \max_T \{KL(p(x|T), q(x|T))\} \quad (4.18)$$

Le mélange de  $K$  Meilleurs Arbres Recouvrants, tel qu'il est décrit aux équations (4.13) - (4.15), sera dit plus performant que le modèle donné par un Meilleur Arbre Recouvrant si et seulement si les distributions de probabilités résultantes du mélange approximent en mieux les probabilités réelles vis-à-vis d'un couple de probabilités  $(p(x|T^{op}), q(x|T^{op}))$  quelconque modélisé par l'un des arbres considérés. Une traduction formelle de ceci sera :

$$KL(p(x|T^{op}), q(x|T^{op})) \leq KL(p(x|T_{mix}^{\Theta^K}), q(x|T_{mix}^{\Theta^K})) \quad (4.19)$$

Afin de justifier le choix de mélange des Meilleurs Arbres Recouvrants, il faut vérifier la validité de l'inégalité (4.19).

Soit  $T_{j_0}$  un Meilleur Arbre Recouvrant quelconque. Comparons la divergence de Kullback - Leibeler entre  $p(x|T_{mix}^{\Theta^K})$  et  $q(x|T_{mix}^{\Theta^K})$  à celle entre  $p(x|T_{j_0})$  et  $q(x|T_{j_0})$ , nous avons :

$$KL(p(x|T_{mix}^{\Theta^K}), q(x|T_{mix}^{\Theta^K})) = \sum_{i=1}^k p(x_i|T_{mix}^{\Theta^K}) \log \left( \frac{p(x_i|T_{mix}^{\Theta^K})}{q(x_i|T_{mix}^{\Theta^K})} \right) \quad (4.20)$$

$$= \sum_{i=1}^N \left( \sum_{j=1}^k \lambda_j p(x_i|T_j) \right) \log \left( \frac{\sum_{j=1}^k \lambda_j p(x_i|T_j)}{\sum_{j=1}^k \lambda_j q(x_i|T_j)} \right) \quad (4.21)$$

Puisque les termes  $\lambda_j p(x_i|T_j)$  et  $\lambda_j q(x_i|T_j)$  sont non-négatifs ( $i = 1, \dots, N; j = 1, \dots, k$ ), selon le théorème de l'inégalité du Log-somme (*Log-sum-inequality*), nous obtenons :

$$\begin{aligned} \sum_{i=1}^N \left( \left( \sum_{j=1}^k \lambda_j p(x_i|T_j) \right) \log \left( \frac{\left( \sum_{j=1}^k \lambda_j p(x_i|T_j) \right)}{\left( \sum_{j=1}^k \lambda_j q(x_i|T_j) \right)} \right) \right) \\ \leq \sum_{i=1}^N \left( \sum_{j=1}^k \lambda_j p(x_i|T_j) \log \frac{\lambda_j p(x_i|T_j)}{\lambda_j q(x_i|T_j)} \right) \end{aligned} \quad (4.22)$$

$$= \sum_{i=1}^N \left( \sum_{j=1}^k \lambda_j p(x_i|T_j) \log \frac{p(x_i|T_j)}{q(x_i|T_j)} \right) \quad (4.23)$$

$$= \sum_{j=1}^k \left( \lambda_j \sum_{i=1}^N p(x_i|T_j) \log \frac{p(x_i|T_j)}{q(x_i|T_j)} \right) \quad (4.24)$$

$$= \sum_{j=1}^k (\lambda_j KL(p(x|T_j), q(x|T_j))) \quad (4.25)$$

Soit  $j_0 \in \{1, 2, \dots, k\}$ , par définition nous avons :

$$\forall j \in \{1, 2, \dots, k\} \quad KL(p(x|T_{j_0}), q(x|T_{j_0})) = KL(p(x|T_j), q(x|T_j)) \quad (4.26)$$

donc

$$KL(p(x|T_{mix}^{\Theta^K}), q(x|T_{mix}^{\Theta^K})) \leq \sum_{j=1}^k (\lambda_j KL(p(x|T_{j_0}), q(x|T_{j_0}))) \quad (4.27)$$

d'où l'inégalité

$$KL(p(x|T_{mix}^{\Theta^K}), q(x|T_{mix}^{\Theta^K})) \leq KL(p(x|T_{j_0}), q(x|T_{j_0})) \quad (4.28)$$

Du moment que l'objectif des Meilleurs Arbres Recouvrants est de maximiser la divergence entre les deux probabilités approximées, nous déduisons que le mélange d'arbres (équation (4.4)) n'offrira pas de meilleurs résultats par rapport au modèle fondé sur un Meilleur Arbre Recouvrant. Par suite, en guise d'utiliser un mélange dans le cas des Meilleurs Arbres Recouvrants, il est recommandé de procéder à une autre alternative pour la quête de l'arbre idéal parmi ce genre d'arbres.

### 4.2.3.3 Mélange des arbres Discriminatoires

L'arbre de dépendances à bi-probabilité peau et non peau Discriminatoire a été proposé au chapitre précédent dans le cadre de la recherche d'une structure qui permettra d'améliorer l'ensemble des arbres aléatoires modélisant les distributions de probabilités peau et non peau. Le critère d'optimisation proposé pour cet arbre est le suivant :

$$\tilde{T} = \arg \min_T \{KL(p(x), p(x|T)) + KL(q(x), q(x|T)) - KL(p(x|T), q(x|T))\} \quad (4.29)$$

Cet arbre permet à la fois de gérer explicitement et l'interclasse et l'intra classe, susceptibles d'exister entre les deux classes peau et non peau dont les distributions de probabilités sont approximées par  $p(\cdot|T)$  et  $q(\cdot|T)$ . Or, lorsqu'une multiplicité d'arbres de ce genre existe sur un même graphe, il vaut mieux exploiter d'avantage les différents apports de tous les arbres possibles. Pour juger que le mélange proposé permettra de définir l'arbre recouvrant "idéal" performant les probabilités modélisées par tous les arbres Discriminatoires possibles, la condition nécessaire et suffisante est que le couple de probabilité  $(p(\cdot|T_{mix}^{\Theta^K}), q(\cdot|T_{mix}^{\Theta^K}))$  modélisé par l'arbre du mélange minimise les distances d'intra classe et maximise celles d'interclasse, mieux que le couple  $(p(\cdot|T^{op}), q(\cdot|T^{op}))$  modélisé par un arbre Discriminatoire aléatoire. Formellement, nous écrivons :

$$\begin{aligned} & \overbrace{\{KL(p(x), p(x|T_{mix}^{\Theta^K})) + KL(q(x), q(x|T_{mix}^{\Theta^K})) - KL(p(x|T_{mix}^{\Theta^K}), q(x|T_{mix}^{\Theta^K}))\}}^A \\ & \leq \underbrace{\{KL(p(x), p(x|T^{op})) + KL(q(x), q(x|T^{op})) - KL(p(x|T^{op}), q(x|T^{op}))\}}_B \end{aligned} \quad (4.30)$$

Le critère d'optimisation de l'arbre Discriminatoire est une combinaison de critères de l'arbre de Chow Liu et celui du Meilleur Arbre Recouvrant. Suite aux calculs présentés dans les sections précédentes, nous constatons que :

$$KL(p(x|T_{mix}^{\Theta^K}), q(x|T_{mix}^{\Theta^K})) \leq KL(p(x|T^{op}), q(x|T^{op})) \quad (4.31)$$

$$KL(p(x), p(x|T_{mix}^{\Theta^K})) \leq KL(p(x), p(x|T^{op})) \quad (4.32)$$

$$KL(q(x), q(x|T_{mix}^{\Theta^K})) \leq KL(q(x), q(x|T^{op})) \quad (4.33)$$

Ainsi, nous concluons que nous ne pouvons pas décider laquelle des deux mesures  $A$  et  $B$  est la plus grande. Par conséquent, les performances du mélange des arbres Discriminatoires ne peuvent pas être déterminées théoriquement, ce n'est que le domaine et les données d'application utilisant ses arbres qui peuvent les juger.

Nous avons montré que le mélange des arbres recouvrants optimaux, tel qu'il est exprimé par l'équation (4.4), ne permet pas d'améliorer les résultats obtenus par les Meilleurs Arbres Recouvrants possibles indépendamment considérés. Toutefois, en ce qui concerne les arbres à bi-probabilité peau et non peau Discriminatoires, les performances offertes par le mélange ne peuvent pas être prédites théoriquement. Afin d'exploiter les connaissances apportées par les deux types d'arbres, il faut adapter un autre procédé de combinaison.

### 4.3 Mélange des arbres

Dans les quatre études théoriques présentées ci-dessus, l'arbre de mélange  $T_{mix}$  est supposé être déterminé pour chacun des modèles étudiés. La présente section a pour fin de décrire le modèle de mélange conjoint, et des arbres et de leurs distributions de probabilités modélisées, ainsi que la procédure adoptée pour la définition de l'arbre résultant du mélange.

En se basant sur la théorie mathématique, il semble beaucoup plus approprié d'utiliser le mélange des arbres recouvrants optimaux dans le cas des arbres de Chow Liu ou encore dans le cas des arbres de dépendances à bi-probabilité peau et non peau Simples. Nous considérons le cas particulier des arbres de Chow Liu appliqués à la détection de la peau pour présenter le modèle du mélange. Bien entendu, pour déterminer le modèle du mélange des arbres de dépendances à bi-probabilité Simples il suffit de substituer les coûts des arcs du modèle que nous allons présenter.

### 4.3.1 Modèle de mélange

Le cardinal de l'ensemble des arbres de dépendances ou encore de recouvrements optimaux sur un graphe est susceptible d'être supérieur à un. Sous réserve qu'aucune contrainte ou caractéristique supplémentaire sur les différents arbres n'est additionnée, pour sélectionner la structure de dépendances la plus adéquate à l'indexation d'un modèle probabiliste, la méthode de mélange des `ArbreROpts` est proposée. En fait, cette méthode détermine l'arbre recouvrant idéal (arbre du mélange) modélisant au mieux une ou deux distributions de probabilités en fonction du graphe considéré.

Cependant, pour définir un arbre de mélange des arbres recouvrants optimaux, il est avantageux d'exploiter les connaissances à priori. Dans notre cas, ces connaissances sont les arbres recouvrants optimaux qui encapsulent un tas d'importantes dépendances et évidences entre les données. D'ailleurs, nous sommes convaincus que l'arbre idéal en quête est au coeur des arbres de dépendances optimaux. D'où la naissance de l'idée de mélange des structures des `ArbreROpts`. La Figure 4.2 illustre le principe global de la définition de l'arbre du mélange et par suite l'estimation de la probabilité du mélange.



FIGURE 4.2 – Procédé du mélange des  $K$  arbres recouvrants optimaux.

Dans la chaîne présentée à la Figure 4.2, le seul maillon inconnu pour le moment est le bloc du "modèle du graphe de mélange". Ce dernier constitue le noyau du modèle du mélange des `ArbreROpts`. Pour définir avec rigueur ce bloc, nous décomposons la construction du graphe de mélange en deux étapes :

– Construction du graphe de mélange non pondéré :

Chacun des arbres de dépendances optimaux possibles emmagasine les dépendances et les interactions importantes entre les données. La suppression d'un arc de l'un de ces

arbres engendre une perte de l'information pertinente fournie par le critère d'optimisation sur lequel est basé le modèle d'arbres en question. D'autant plus que le modèle du graphe de mélange proposé doit assurer la restauration des différents *ArbreROpts*. Autrement dit, pour un arbre  $T$  appartenant à l'ensemble des *ArbreROpts*  $\Theta^K$ , si pour tout arbre  $T'$  dans  $\Theta^K$  différent de  $T$  le coefficient du mélange  $\lambda(T')$  est nul alors l'arbre du mélange  $T_{mix}$  n'est autre que l'arbre  $T$ . Il apparaît alors astucieux de conserver toutes les dépendances apportées par les différents *ArbreROpts*.

Du moment que tous les *ArbreROpts* sont définis à partir du même ensemble de noeuds et ne diffèrent que sur l'ensemble des arcs, nous pouvons définir le graphe du mélange  $G_{mix}(N_{mix}, A_{mix})$  comme suit :

- \* L'ensemble des noeuds  $N_{mix}$  est le même que l'ensemble des noeuds d'un *arbreROpt* quelconque  $N$  ;
- \* Les relations entre les noeuds sont définies par l'ensemble des arcs  $A_{mix} = \bigcup_{T \in \Theta^K} A_T$ , avec  $A_T$  est l'ensemble des arcs d'un arbre  $T$ .

La Figure 4.3 illustre un exemple de construction de du modèle du graphe non pondéré du mélange.

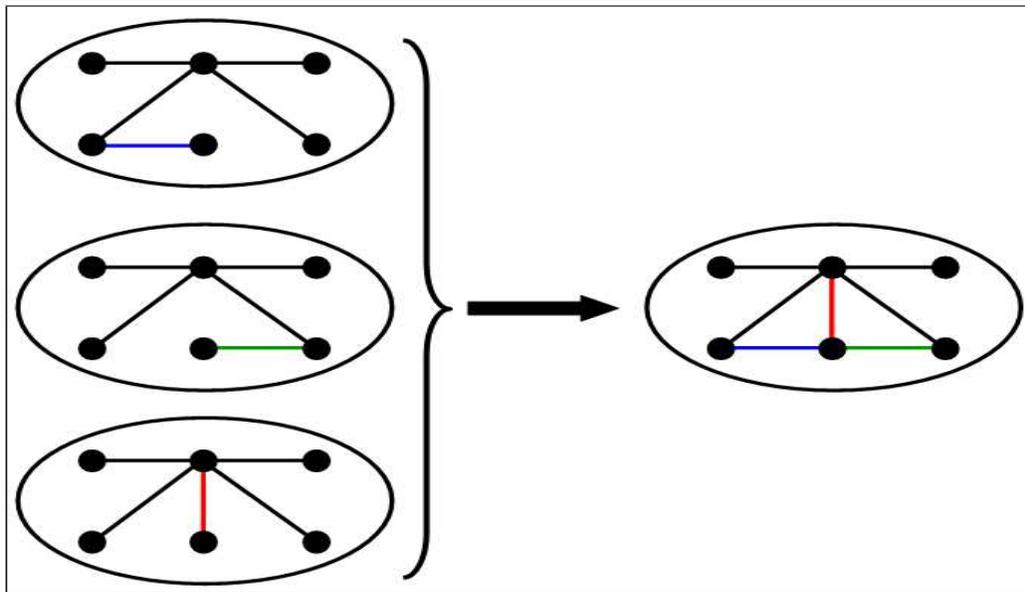


FIGURE 4.3 – Exemple de construction du graphe du mélange à partir de trois arbres de recouvrement ou de dépendances optimaux.

– Pondération du graphe de mélange :

Après avoir déterminé le graphe non pondéré du mélange, il est nécessaire de déterminer les poids de ses différents arcs pour obtenir le graphe pondéré du mélange. Pour cela, nous proposons de définir le coût d'un arc reliant deux noeuds voisins  $u$  et  $v$  par :

$$W_{u \sim v} = KL(Pmix_{uv}(x_u, x_v), Pmix_u(x_u)Pmix_v(x_v)) \quad (4.34)$$

avec  $Pmix_u$  est une probabilité marginale de la probabilité du mélange  $Pmix$  au niveau d'un noeud, et  $Pmix_{uv}$  sa distribution marginale au niveau de chaque couple de noeuds. Aussi la probabilité du mélange  $Pmix$  conditionnée par l'arbre de mélange  $T_{mix}$  s'écrit sous la forme :

$$Pmix(x|T_{mix}) = P(x|T_{mix}) \quad (4.35)$$

avec  $P$  est la probabilité réelle.

A ce niveau, le modèle du graphe de mélange est achevé. Dans la perspective de déterminer l'arbre du mélange, il est demandé de décrire la manière avec laquelle les paramètres de ce mélange seront estimés. Ceci est l'objectif de la section ci-dessous.

### 4.3.2 Estimations de paramètres du modèle de mélange

Nous utilisons une variante de la classe d'algorithme Espérance-Maximisation (EM) pour trouver le maximum de vraisemblance des paramètres du modèle probabiliste du mélange. Notons le  $\Phi = (\lambda(T)^1, \lambda(T)^2, \dots, \lambda(T)^K)$  le paramètre du mélange des probabilités données par les arbres recouvrants optimaux. Ainsi, la log-vraisemblance du paramètre  $\Phi$  est donnée par

$$L(x, \Phi) = \sum_{i=1}^n \log \left( \sum_{j=1}^K \lambda(T)^j P(x_i|T^j) \right) \quad (4.36)$$

Dans l'objectif de résoudre cette équation, nous proposons une instance de l'algorithme EM. Cette instance est décrite comme suit :

– Etape E : utilisation de la règle de Bayes pour le calcul de :

$$\gamma_j^i = \frac{\lambda(T)^j P(x_i|T^j)}{\sum_l \lambda(T)^l P(x_i|T^l)}. \quad (4.37)$$

- Etape M : détermination des coefficients du mélange minimisant la divergence de Kullback entre la probabilité réelle et celle estimée.
- L’algorithme converge lorsque la différence entre, la divergence KL entre la probabilité réelle et la probabilité  $P^N(x|T_{mix})$  d’une part, et la divergence KL entre la probabilité réelle et la probabilité  $P^{(N-1)}(x|T_{mix})$  de l’autre part, est négligeable.  
Avec  $P^{(N)}(x|T_{mix})$  et  $P^{(N-1)}(x|T_{mix})$  sont respectivement les probabilités estimées à l’itération  $N^{th}$  et  $(N-1)^{th}$  de l’algorithme ;  $T_{mix}$  étant l’arbre du mélange déterminé à chaque itération.

A noter que la condition de convergence que nous proposons dans cet algorithme est déduite par le modèle de mélange des structures et des probabilités des arbres de dépendances optimaux.

La procédure suivante synthétise les démarches définissant comment les coefficients du mélange des arbres de dépendances optimaux ainsi que l’arbre de mélange seront déterminés.

**Procédure 2.** *Mélange de  $K$  Arbres de dépendances optimaux*

- Entrée : [ $K$  Arbres de dépendances optimaux et le modèle initial  $\{k, \lambda(T)^j, j = 1, 2, \dots, k\}$ ].

Calculer les paramètres :

$$\gamma_j^i = \frac{\lambda(T)^j P(x_i|T^j)}{\sum_l \lambda(T)^l P(x_i|T^l)}.$$

$$\Gamma^j = \sum_{i=1}^n \gamma_j^i \quad , \quad j = 1, 2, \dots, k$$

Dresser le graphe pondéré du mélange des  $k$  arbres de dépendances optimaux.

Itérer jusqu’à convergence :

+ Etape E :

pour  $j = 1, 2, \dots, k$  ,  $i = 1, 2, \dots, n$ , calculer  $\gamma_j^i$ .

+ Etape M :

pour  $j = 1, 2, \dots, k$  ,  $\lambda(T)^j \leftarrow \frac{\Gamma^j}{n}$

calculer le coût de l’arc reliant deux noeuds  $u$  et  $v$  et défini par

$$KL(Pmix_{uv}(x_u, x_v), Pmix_u(x_u)Pmix_v(x_v))$$

Utiliser l’algorithme de Kruskal pour déterminer l’arbre recouvrant optimal.

- Sortie : [ Arbre du mélange  $T_{mix}$ , ainsi que le modèle final  $\{k, \lambda(T)^j, j = 1, 2, \dots, k\}$ ].

### 4.3.3 Quelques résultats expérimentaux

Nous présentons ici quelques résultats expérimentaux issus de la détection de la peau dans les images en couleurs en utilisant le modèle du mélange des arbres proposé. Nous considérons en particulier des arbres de type Chow Liu possibles. Précisément, nous considérons deux arbres de ce type. La Figure 4.4 illustre quelques résultats obtenus.

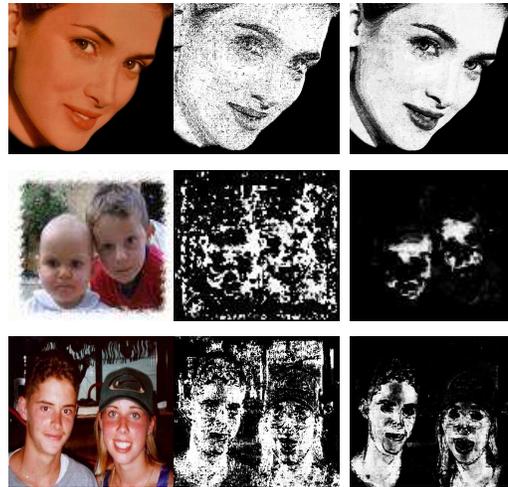


FIGURE 4.4 – Quelques entrées et sorties des modèles de mélange des arbres pour la détection de la peau. La deuxième colonne représente les images aux niveaux de gris du modèle où un seul arbre est considéré alors que la troisième colonne donne celles du modèle où deux arbres sont mélangés.

La Figure 4.4 représente quelques images couleurs (entrées des modèles) et leurs résultats obtenus par un modèle de peau fondé sur un arbre de Chow Liu et un deuxième modèle basé sur le mélange de deux arbres possibles vérifiant le critère d'optimisation de l'algorithme de Chow Liu. Nous attirons l'attention que l'arbre de Chow Liu utilisé dans le modèle considérant un seul arbre, est celui obtenu aléatoirement par l'algorithme de Chow Liu. Nous constatons que la qualité de la détection de la peau est nettement meilleure sur les sorties du modèle de mélange de deux arbres. Cette qualité est obtenue grâce à la considération de deux arbres à mélanger à la place d'un seul arbre. Cela montre que l'arbre ajouté emmagasine des informations importantes pour la séparation des pixels peau de ceux non peau, et qui ont été ignorées par le modèle fondé sur un seul arbre. Ainsi, nous concluons qu'en plus du fondement mathématique présenté précédemment,

l'expérimentation prouve, elle aussi, que le mélange des arbres possibles vérifiant le critère d'optimisation de l'arbre de Chow Liu présente les meilleures performances.

## 4.4 Pour une combinaison des arbres recouvrants optimaux

Comme nous l'avons déjà cité précédemment, les apports du mélange aussi bien dans le cas des arbres de dépendances à bi-probabilité peau et non peau Discriminatoires que dans le cas des Meilleurs Arbres Recouvrants, ne peuvent pas être prédits. En fait, pour chacun de ces deux catégories d'arbres, l'étude théorique menée sur le mélange des `ArbreROpts` a prouvé qu'il est incertain que l'arbre du mélange soit l'arbre idéal. D'autant plus, l'arbre de mélange peut bien avoir des performances dégradées que celles offertes par le modèle basé sur un arbre. En tenant compte de ce constat, il faut procéder à une autre alternative que le mélange pour déterminer au mieux l'arbre idéal.

Dans ce contexte, la combinaison des différents arbres de dépendances optimaux semble plus appropriée pour obtenir l'arbre optimal idéal. Dans cette section, nous présenterons un panorama des méthodes de combinaisons existantes, ensuite nous décrirons notre approche de combinaison pour un arbre recouvrant idéal.

### 4.4.1 Méthodes de combinaison

Un classifieur n'est pas uniquement un dispositif qui se limite à donner le type d'un objet inconnu mais c'est aussi une source d'information. L'objectif principal de la combinaison de classifieurs est l'amélioration de la qualité de reconnaissance du système par rapport à chacun des classifieurs isolés, et ce par la collecte de l'information issue de plusieurs sources. Dans ce contexte, plusieurs topologies de combinaisons de classifieurs sont envisageables [101][102]. Elles peuvent être subdivisées en quatre catégories :

- Topologie séquentielle. Elle est aussi dite combinaisons en série. Elle consiste à placer un classifieur à la sortie d'un autre. Ainsi à chaque étape, l'entrée d'un classifieur est composée de la sortie du classifieur lui précédant. Quant à sa sortie, elle constitue l'entrée du classifieur lui succédant. Il est donc clair que dans ce type de combinaison, la décision finale est influencée par l'ordre dans lequel les classifieurs sont alignés.

- Topologie parallèle. Dans ce cas de figure, les classifieurs sont supposés être indépendants. En effet, chaque classifieur opère comme s'il existait seul du moment que son entrée ne dépend pas des autres classifieurs. La combinaison est réalisée sur toutes les sorties des classifieurs au même niveau.
- Topologie hybride. La combinaison hybride est elle même une combinaison d'un ensemble de classifieurs placés en série et en parallèle les uns par rapport aux autres. L'objectif derrière une telle topologie réside dans la définition d'un procédé permettant de déterminer la décision finale la plus certaine. Or, ce type de combinaison est coûteux lorsque le nombre de classifieur à combiner est élevé. Cela est dû au fait qu'il faut tester toutes les combinaisons possibles.
- Autres topologies. Cette catégorie englobe des topologies de combinaisons utilisant des boucles sur un ou plusieurs classifieurs, celles basées sur des interactions entre les classifieurs, ou encore d'autres utilisant et l'entrée et la sortie d'un classifieur dans la prise de décision, ...

Bien entendu, pour pouvoir combiner un ensemble de classifieurs, il est nécessaire qu'ils opèrent de la même manière. Autrement dit, il faut que leurs entrées et leurs sorties soient homogènes.

La deuxième topologie, à savoir la topologie parallèle, est la plus répandue. En fait, celle-ci est basée sur le concept de démocratie vis-à-vis des différents classifieurs à combiner. Ce concept est le plus adapté lorsqu'une connaissance à priori sur les classifieurs est indisponible. Dans ce contexte, nous donnons quelques méthodes des combinaisons parallèles. En particulier, nous nous limitant aux combinaisons des classifieurs dont les sorties sont de type classe ou mesure. Nous rappelons que les classifieurs probabilistes sont des classifieurs de mesure mais aussi ils sont une variante des classifieurs de classe. En effet, ils affectent à une classe la mesure de la confiance ou de l'incertitude accordée à la réponse proposée.

Les méthodes de combinaisons parallèles de classe ou mesure, sont de deux types : non paramétriques et paramétriques.

- Les méthodes non paramétriques. Dans le cas des classifieurs de classe, ce type de méthodes se restreint principalement aux combinaisons par vote. Plusieurs modèles du vote existent tels que le vote à la majorité, le vote à la pluralité, le vote notoire, etc. Quant aux combinaisons de mesure, ce type de méthodes est basé sur l'utilisation de

règles fixes telles que la somme, le produit, le maximum, le minimum, ou encore sur l'utilisation de la moyenne de Bayes.

- Les méthodes paramétriques. Lorsqu'il s'agit de classifieurs de type classe, il y a souvent recours au vote avec pondération, à la méthode de Bayes ou encore aux méthodes basées sur la théorie de l'évidence. Concernant les classifieurs de mesure, ils adoptent le plus souvent des méta-classifieurs tels que les réseaux de neurones.

Dans la conception de notre système de coopération des classifieurs fondés sur les ArbresROpts, si nous optons pour une combinaison parallèle plusieurs méthodes de celles présentées ci-dessus seront à éliminer. En effet, les études théoriques sur le mélange des ArbresROpts Discriminatoires et des Meilleurs Arbres Recouvrants, ont désapprouvé l'utilisation du mélange. Or, une méthode de combinaison non paramétrique utilisant les sommes des mesures ou encore une combinaison paramétrique de genre vote avec pondération, sont des variantes du mélange.

#### 4.4.2 Approche pour la combinaison des arbres recouvrants optimaux

Avant d'exposer notre proposition pour la combinaison des ArbresROpts, nous attirons l'attention que l'ensemble des Meilleurs Arbres Recouvrants et/ou l'ensemble des arbres de dépendances à bi-probabilité peau et non peau Discriminatoires seront référencés par l'ensemble des arbres de dépendances optimaux.

Nous supposons que l'ensemble  $\Theta(G)^k$  des arbres de dépendances optimaux indexant les probabilités peau et non peau, est déjà déterminé à la phase d'apprentissage. La tâche suivante consiste à définir une approche de combinaison préservant et exploitant les informations apportées par les diverses sources de probabilités, et cela afin d'améliorer la prise de décision lors du calcul des inférences.

Considérons  $K$  classifieurs probabilistes dont chacun est fondé sur un arbre de dépendances optimal. Etant donné que nous ne procédons pas d'autres informations pour privilégier et déterminer la capacité individuelle d'un arbre par rapport aux autres, il est plus adéquat de traiter les différents classifieurs à pied d'égalité. De là, les différents classifieurs doivent coopérer dans une stratégie issue des combinaisons parallèles. Ainsi, nous présentons une approche de combinaison subdivisée en trois étapes consécutives :

1. Sortie d'un classifieur :

Pour chaque classifieur, calculer la probabilité  $P(y_s|x, T_j)$  au niveau d'un pixel  $s$  sachant des informations sur ses voisins constituant le vecteur  $x$ ; avec  $T_j$  ( $j = 1, 2, \dots, k$ ) est l'arbre sur lequel est fondée le  $j^{eme}$  classifieur.

2. Ensembles de classes :

Notons  $C_1$  et  $C_2$  respectivement les deux classes peau et non peau.

Comparer chaque probabilité  $P(y|x, T_j)$  à un seuil  $s_1$ . Ce dernier paramètre permet de déterminer la probabilité d'appartenance d'un vecteur  $x$  à la classe  $C_1$ ; il est calculé dans la phase d'apprentissage, lors de la construction des histogrammes correspondants aux classes  $C_1$  et  $C_2$ .

La comparaison permet de partager les différentes probabilités calculées en deux ensembles : (1) ensemble  $E_1$  correspondant à la classe  $C_1$  et (2) ensemble  $E_2$  correspondant à la classe  $C_2$ . Formellement, les deux ensembles  $E_1$  et  $E_2$  sont définis par

$$E_1 = \{P(y|x, T_j) \ / \ P(y|x, T_j) \geq s_1\} \quad (4.38)$$

$$E_2 = \{P(y|x, T_j) \ / \ P(y|x, T_j) < s_1\} \quad (4.39)$$

3. Sortie du modèle de la combinaison :

Dans un premier temps, utiliser un algorithme de vote pour déterminer la classe dominante. Nous appelons classe dominante la classe qui reçoit plus de vote. Le vote utilisé consiste à comparer la probabilité en sortie d'un classifieur par rapport au seuil  $s_1$ . Si cette probabilité est supérieure ou égale à  $s_1$  c'est la classe  $C_1$  qui reçoit le vote, dans le cas opposé c'est la classe  $C_2$  qui reçoit le vote.

Le plus grand cardinal des deux ensembles  $E_1$  et  $E_2$  permet de déterminer la classe dominante, soit  $C_d$  (égale à  $C_1$  ou  $C_2$ ).

Dans un deuxième temps, nous calculons  $P(y|x)$  la probabilité globale du modèle coopératif. Nous choisirons cette probabilité égale au maximum des probabilités désignant la classe  $C_d$  dans le cas où  $C_d = C_1$ . Dans le cas où  $C_d = C_2$ , la probabilité  $P(y|x)$  sera égale au minimum des probabilités désignant la classe  $C_2$ . Formellement,

nous écrivons :

$$\text{Si } |E_1| \geq |E_2| \quad \text{Alors} \quad P(y|x) = \max\{E_1\} \quad (4.40)$$

$$\text{si } |E_1| < |E_2| \quad \text{Alors} \quad P(y|x) = \min\{E_2\} \quad (4.41)$$

L'idée motrice derrière cette combinaison se résume dans la perception d'un modèle capable de restaurer chaque modèle d'arbre et par la même occasion n'introduire aucune contrainte supplémentaire sur les différents classifieurs. D'ailleurs, via la méthode de combinaison que nous proposons, nous garantissons qu'à chaque fois, la probabilité en sortie du modèle coopératif n'est autre que la probabilité approximée par l'un des *ArbreROpt*. Cette exigence reste conforme avec la définition des *ArbreROpt*, à savoir que chaque *ArbreROpt* modélise la distribution de probabilité optimale. Ainsi, aucune autre mesure en dehors de celles données par les différents classifieurs ne peut déterminer la meilleure estimation de la probabilité globale calculée.

En considérant deux modèles de peau fondés sur les Meilleurs Arbres Recouvrants possibles, la Figure 4.5 illustre quelques entrées (images couleurs) et sorties (images aux niveaux de gris) de ces deux détecteurs de peau.



FIGURE 4.5 – Quelques entrées et sorties des modèles fondés sur la combinaison des arbres. Images sur la deuxième colonne correspondent au modèle du Meilleur Arbre Recouvrant alors que les images de la troisième colonne correspondent à celui de la combinaison de trois Meilleurs Arbres Recouvrants possibles.

Les images aux niveaux de gris correspondent aux sorties du détecteur de peau basé sur le modèle d'un Meilleur Arbre Recouvrant (deuxième colonne de la Figure 4.5) et

celui fondé sur la combinaison de deux Meilleurs Arbres Recouvrants possibles (troisième colonne de la Figure 4.5). Nous remarquons que la bonne détection de pixels peau et ceux non peau est obtenue par le modèle de la combinaison des arbres considérés. Cela justifie expérimentalement que la combinaison des arbres améliore la qualité de la détection de la peau dans les images couleurs.

## 4.5 Discussion

L'arbre de recouvrement optimal (ArbreROpt) a trouvé son utilisation dans un grand ensemble de domaines dont les applications utilisent des modèles graphiques et invoquent le principe d'optimisation. Toutefois, sur un même graphe, il est possible que plusieurs ArbresROpts coexistent. Cette coexistence est traduite par la problématique de multiplicité d'ArbresROpts. Nous Parlons de problématique, pour la simple raison que dans le cas où aucune information supplémentaire n'est fournie, tous les arbres sont considérés à pieds d'égalité. Néanmoins, chaque ArbreROpt détermine une structure de dépendances bien particulière entre les données, et par la suite une estimation particulière de la probabilité réelle. Or, ce qui est souvent exigé par les systèmes ce n'est pas un ArbreROpt mais plutôt l'arbre idéal modélisant la distribution de probabilité optimale estimée. Cet arbre est dit idéal car en plus du fait qu'il répond au critère d'optimisation désignant l'ensemble des ArbresROpts, il encapsule des propriétés qui ne sont pas aussi faciles à quantifier.

En outre, à cause de la discontinuité des données manipulées, il est nécessaire de tenir compte de facteurs pouvant ajuster les calculs. Dans ce contexte, nous avons proposé d'utiliser une " erreur de tolérance " qui désigne la marge d'erreur que l'utilisateur peut accepter. Un tel paramètre offre la possibilité d'élargir l'ensemble des ArbresROpts sélectionné. Par ce principe, l'appartenance de l'arbre idéal à ce dernier ensemble est garantie. En effet, l'erreur de tolérance ne permet pas seulement d'ajuster les calculs mais aussi c'est une manière de quantifier des caractéristiques implicites de la distribution à indexer.

Le manque de connaissance sur l'arbre idéal à sélectionner parmi un ensemble d'ArbresROpts possibles, rend la détermination de cet arbre difficile ou encore impossible. Les algorithmes classiques tels que Kruskal ou Prim ne sont pas déterminants. Autrement dit, lorsqu'un choix entre les différents ArbresROpts se manifeste, ces algorithmes ne possèdent aucune règle qui permet de favoriser un arbre par rapport aux autres. Ainsi l'ArbreROpt

donné dans ce cas est aléatoire. Or, le choix au hasard d'un arbre ne garantit pas que l'arbre obtenu soit l'idéal à considérer.

Nous avons proposé dans le présent chapitre deux alternatives pour la quête de l'arbre idéal. La première approche est un mélange des arbres alors que la deuxième est une combinaison de ceux-ci. Dans les deux cas, notre objectif consiste à élaborer un modèle de peau idéal. Un modèle qui permet d'exploiter les informations apportés par chacun des précédents arbres. Ces deux approches mettent en évidence l'intérêt d'exploiter les connaissances sur les différents arbres de dépendances optimaux possibles sans introduire des informations supplémentaires.

Le mélange de probabilité a prouvé son utilité dans plusieurs domaines en particulier la détection de la peau. Nous citons à titre d'exemple le mélange de gaussienne qui présente des meilleurs résultats vis-à-vis d'une simple gaussienne. De ce principe est naît l'idée de mélange des distributions de probabilités approximées par les différents ArbreROpts. Toutefois, les distributions d'arbres modélisent au mieux les probabilités réelles. Nous avons proposé que l'arbre idéal soit l'arbre modélisant la probabilité du mélange. Ainsi, notre proposition de mélange consiste à mélanger les probabilités modélisées par les différents ArbreROpts ainsi que les structures de ces arbres, conjointement. La procédure du mélange telle qu'elle est présentée, assure la reconstruction de chaque ArbreROpt possible. Des démonstrations théoriques ont été présentées dans ce chapitre afin de prouver l'apport d'une approche fondée sur le mélange par rapport à celle où un ArbreROpt aléatoire est considéré. Nous avons pu constater que théoriquement le mélange proposé permet d'offrir des améliorations dans le cas des ArbreROpts de type arbre de Chow Liu ou encore dans le cas de ceux de type arbre de dépendances à bi-probabilité Simple. De plus, il est théoriquement précisé que ce n'est que le mélange des ArbreROpts qui garantit de présenter de meilleures performances. Cependant, dans le cas des ArbreROpt de type Meilleur Arbre Recouvrant et arbre de dépendance à bi-probabilité Discriminatoire, l'évaluation théorique des performances du modèle de la peau basé sur le mélange de ces arbres n'est pas déterminante. Autrement dit, ce n'est que les résultats expérimentaux qui permettront une telle évaluation. Pour remédier à cela, nous avons proposé la combinaison de ces deux derniers types d'arbres. Du moment qu'aucune connaissance ne permet de différencier entre ces deux derniers modèles d'ArbreROpts, la combinaison d'arbres et de probabilités proposé, est une approche parallèle qui assure de traiter chaque ArbreROpt

à pied d'égalité avec ses semblables.

Théoriquement, il est prouvé que le mélange et la combinaison des probabilités modélisées par les `ArbreROpts` ainsi que des structures de ces arbres, permettent de définir des modèles de peau robustes. Il suffit de choisir pour chaque méthode considérée, l'ensemble des `ArbreROpts` le mieux adapté. Dans le chapitre qui suit, nous proposerons d'évaluer expérimentalement ces différents modèles de peau.

---

## DÉTECTION DE LA PEAU PAR LES NOUVEAUX MODÈLES D'ARBRES PROBABILISTES : EXPERIMENTATIONS ET RÉSULTATS

Les chapitres 3 et 4 présentent les différents modèles de distributions d'arbres que nous proposons pour un double objectif : l'amélioration des distributions de probabilités indexées par les arbres et l'élaboration d'un modèle robuste de la peau dans les images couleurs, capable de déterminer la classe d'appartenance de chaque pixel de l'image. Les modèles fondés sur des arbres de dépendances à bi-probabilité peau et non peau peuvent être partagés en deux ensembles : (1) modèles fondés sur un arbre et (2) ceux basés sur la combinaison des arbres. Quant au troisième modèle proposé, il détermine l'ensemble composé des modèles à base des mélanges d'arbres quelconques.

Le présent chapitre est consacré à l'évaluation de ces différentes catégories de distributions des arbres.

Le chapitre s'organise comme suit : dans un premier temps, le protocole expérimental utilisé pour l'évaluation des classifieurs des pixels peau et non peau issues des différents modèles des arbres, sera présenté. Dans un deuxième temps, les résultats obtenus par les différents modèles fondés sur un arbre seront exposés, suivis des résultats donnés par les mélanges des arbres et enfin ceux offerts par les combinaisons des arbres. Pour chaque ensemble des arbres discuté, une analyse quantitative et une autre qualitative seront élaborées. Notons aussi que les différentes distributions de prédictions seront aussi évaluées par rapport au classifieur de base de la détection de la peau dans les images, à savoir le modèle dit de base.

## 5.1 Protocole expérimental

La présente section décrit le matériel utilisé (corpus des images, espace de couleurs, etc) pour l'évaluation des modèles d'arbres dans le cadre de la détection de la peau, ainsi que les hypothèses et les contraintes adoptées pour la réalisation des expérimentations sur les images couleurs.

### 5.1.1 Corpus

L'un des principaux facteurs qui influence les résultats obtenus par un système, est le corpus utilisé. Ce dernier détermine la base des images d'apprentissage et celle de test, qui permettent de définir et d'évaluer le support de modélisation des données.

La construction des différents modèles d'arbres est fondée sur l'extraction des connaissances de la base d'apprentissage, il convient alors que cette base soit assez large et contienne des échantillons d'images représentant les différentes races, sexes, et conditions d'éclairage et de luminosité. Se restreindre à un ensemble particulier d'images dont les conditions de prise de vues sont connues à l'avance, comme c'est le cas pour une grande majorité des systèmes de la détection de la peau existants, ne permet pas d'aboutir à un modèle de la peau robuste et générique.

Nous utilisons la base d'images Compaq [52] pour la construction et l'évaluation des différents modèles d'arbres proposés, dans la perspective d'obtenir un meilleur modèle de classification de la peau. Cette base est considérée comme la première plus large base dédiée à la classification de la peau dans les images. Elle est constituée de 4 650 images positives et 8 914 d'images négatives. Les images positives sont les images contenant de la peau alors que les négatives ne contiennent pas de la peau. En plus cette base contient une vérité de terrain définie par les différents masques des images manuellement segmentées (images binaires).

Les images de la base Compaq couvrent les différentes problématiques liées à la détection de la peau. En effet, les images de cette base sont prises du Web et remplissent des conditions de prises de vues non limitées (arrière plan simple et complexe, bruit, contraste, ...) et peuvent contenir une ou plusieurs personnes à la fois. La Figure 5.1 représente quelques images de la base Compaq.

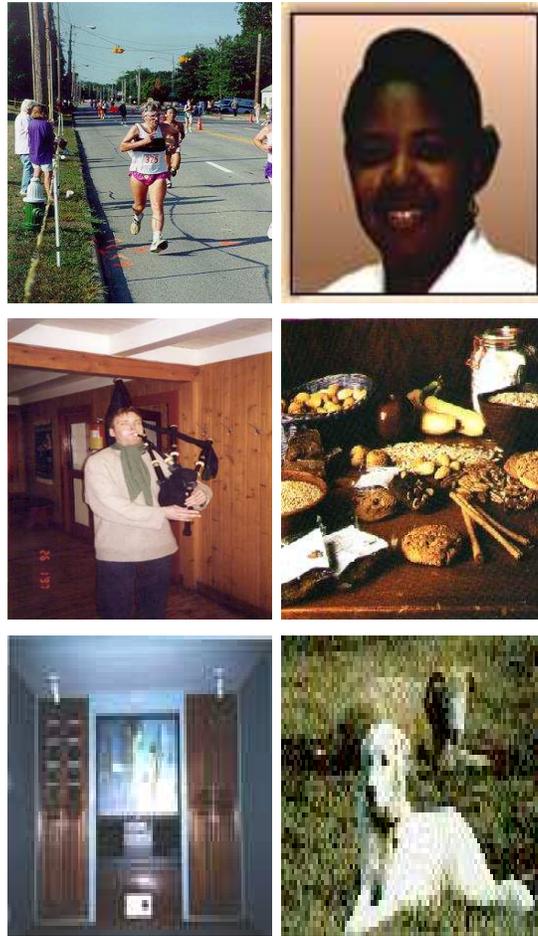


FIGURE 5.1 – Quelques images de la base Compaq.

Afin d'apprendre les structures des différents arbres que nous utilisons par la suite pour définir des détecteurs de la peau, nous considérons une sous base de la base Compaq, constituée de 1 450 images positives et 2 717 images négatives.

### 5.1.2 Espace de couleur

Nous considérons la couleur comme la caractéristique pertinente qui discrimine les pixels peau de ceux non peau dans une image couleur. Pour cela, il est impératif de choisir un espace colorimétrique dans lequel cette caractéristique sera représentée. Bien que l'espace RVB ne soit pas l'espace colorimétrique le plus adapté à la détection de la peau, nous optons pour cet espace pour diverses raisons, parmi lesquelles :

- il est le plus simple à utiliser du fait que les images classiques sont exprimées dans cet espace ;
- d'après le tableau 1.1 présenté au chapitre 1, le modèle bayésien basé sur les histogrammes dans l'espace RVB [86], offre la meilleure performance ;
- le corpus que nous utilisons pour l'évaluation des différents modèles de peau proposés, est assez large.

### 5.1.3 Vecteur des observations

Le système de voisinage d'un pixel  $s$ , que nous utilisons est le suivant :

$$V_s^3 = \{(i, j) / |i - i_s| < 3, |j - j_s| < 3\} \setminus \{(i_s, j_s)\} \quad (5.1)$$

Ainsi pour chaque pixel  $s$ , un voisinage composé d'un bloc  $3 \times 3$  est considéré.

Dans l'espace RVB, chaque pixel est composé de trois composantes R, V et B. Si nous considérons que les noeuds du graphe représentant le vecteur des observations sont définis par les pixels alors le calcul de la loi marginale entre deux noeuds revient au calcul d'une probabilité d'un 6-uplet. Le calcul dans ce cas est lourd, pour le simplifier nous décomposons chaque pixel en ces trois composantes. Il vient donc que la taille du vecteur des observations soit égale à 27 ( $x = (x_1, x_2, \dots, x_{27})$ ).

La Figure 5.2 représente la manière dont les composantes du vecteur des observations considéré sont notées.

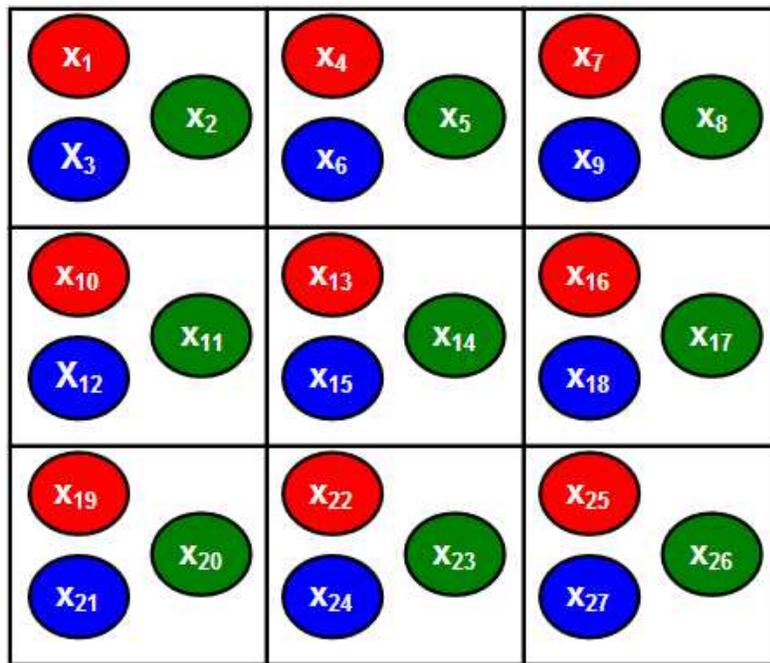


FIGURE 5.2 – Notations utilisées pour les composantes du vecteur des observations issu d'un bloc  $3 \times 3$ .

### 5.1.4 Histogramme

Une manière de calculer empiriquement les probabilités consiste à utiliser les histogrammes. Là aussi nous pouvons justifier le choix de la décomposition des composantes couleurs d'un pixel pour la construction du vecteur des observations. En effet, pour des problèmes de couplage et par suite des tailles des histogrammes à construire (dans RGB, nous avons des histogrammes 3D et 6D), il est plus convenable de considérer les composantes R, V, et B à part. Ainsi, pour un calcul empirique des probabilités marginales de peau et de non peau ( $p_u$ ,  $q_u$ ,  $p_{uv}$  et  $q_{uv}$ ), nous formons des histogrammes 1D et 2D représentant les fréquences d'une ou de deux composantes du vecteur des observations, à la fois.

Un histogramme est caractérisé par un nombre de cases (bin) qui, dans notre cas, quantifie une composante de couleur, et il est aussi caractérisé par les fréquences de répétitions des différentes valeurs que prend cette composante. Les fréquences sont bien entendu calculées à partir de la base d'apprentissage. Quant au nombre de cases plusieurs possibilités se présentent : 32, 64 et 256. Etant donné que les travaux de [52] sur le même corpus (base Compaq) ont penché pour des histogrammes à 32 couleurs, il serait plus convenable de considérer le même nombre de cases.

### 5.1.5 Classifieurs à évaluer

Nous regroupons les classifieurs probabilistes de peau/non peau, présentés dans ce mémoire, selon l'ensemble de types d'arbres qu'ils utilisent. Ainsi nous obtenons trois ensembles, à savoir l'ensemble de classifieurs basés sur :

1. un arbre ;
2. le mélange des arbres ;
3. la combinaison des arbres de dépendances à bi-probabilité.

Nous menons trois études comparatives distinctes sur ces classifieurs :

– Première étude comparative.

Nous considérons les modèles basés sur :

- \* l'arbre de Chow Liu utilisé pour la modélisation de la classe peau et un autre pour la modélisation de la classe non peau ;
- \* l'arbre de dépendances à bi-probabilité peau et non peau Simple ;
- \* le Meilleur Arbre recouvrant ;
- \* l'arbre de dépendance à bi-probabilité Discriminatoire.

– Deuxième étude comparative.

Elle concerne le mélange des arbres optimaux vérifiant le critère d'optimisation de l'arbre de Chow Liu. Pour cela, nous comparons quatre classifieurs fondés sur quatre mélanges différents de ces arbres.

– Troisième étude comparative.

Cette étude est consacrée à la comparaisons de classifieurs basés sur les combinaisons d'arbres de type Meilleur Arbre Recouvrant.

Aussi, afin de pouvoir comparer les performances des différents classifieurs considérés dans les études comparatives que nous menons, nous considérons la même base des images de test dans les différentes études.

La terminologie utilisée dans les différentes études comparatives des détecteurs de la peau, sera présentée au fur et à mesure des études menées.

Nous notons que les résultats qui seront présentés dans les différentes études, sont obtenus par l'analyse directe des sorties des classifieurs et sans l'utilisation des post-traitements tels que la dilatation ou l'érosion.

## 5.2 Première étude comparative

Nous nous intéressons à la comparaison de quelques classifieurs dont chacun est basé sur un arbre. Ainsi nous considérons les modèles basés sur :

- 2-BDTs : modèle considérant les arbres de Chow Liu ;
- diff-BDT : modèle fondé sur le Meilleur Arbre recouvrant ;
- 1-BDT : modèle à base de l'arbre de dépendances à bi-probabilité Simple ;
- 1-BDT-diff-BDT : modèle fondé sur l'arbre de dépendance à bi-probabilité Discriminatoire.

Quelques exemples de résultats obtenus par les différents classifieurs considérés, sont illustrés par la Figure 5.3. L'entrée de chaque classifieur est une image couleur représentée sur la première colonne de la Figure 5.3. Quant à la sortie d'un classifieur, c'est une image au niveau de gris représentée sur les autres colonnes. La troisième colonne donne les résultats obtenus par le classifieur à base de 2-BDTs. La quatrième colonne de la même figure désigne les sorties du classifieur fondé sur 1-BDT. Ensuite, la colonne suivante offre les résultats obtenus par un détecteur de peau fondé sur diff-BDT. Finalement, la sixième colonne représente les sorties d'un modèle de détection basé sur 1-BDT-diff-BDT. En outre, la Figure 5.3 fournit les images aux niveaux de gris obtenues par le classifieur de base qui suppose l'indépendance entre les composantes du vecteur d'observations[52]. Ce dernier classifieur ne prend pas en compte le fait qu'une zone de peau n'est pas aléatoire mais qu'elle constitue plutôt une grande région où des motifs sont répétés [103]. Les sorties de ce classifieur dit de base, sont données sur la deuxième colonne de la figure analysée.

Les images couleurs données par la Figure 5.3 représentent quelques défis de la détection de la peau (changement de luminance et du contraste, arrière plan complexe, des personnes de couleurs de peau différentes, etc). Les différentes sorties des détecteurs sont des images aux niveaux de gris dont la couleur blanche correspond à un pixel peau, la couleur noire correspond à un pixel non peau et les niveaux de couleurs entre les deux couleurs indiquent les degrés de croyance qu'un pixel appartient à la classe peau.

A partir des résultats illustrés par la Figure 5.3, nous constatons que le taux de la détection de la peau s'améliore allant du modèle basé sur 2-BDTs aux trois autres modèles de classifieurs basés sur les arbres de dépendances à bi-probabilité peau et non peau. Cela montre que le choix de représenter et la distribution de la peau et la distribution de non peau sur un même arbre, est justifié expérimentalement. De plus, les classifieurs fondés sur

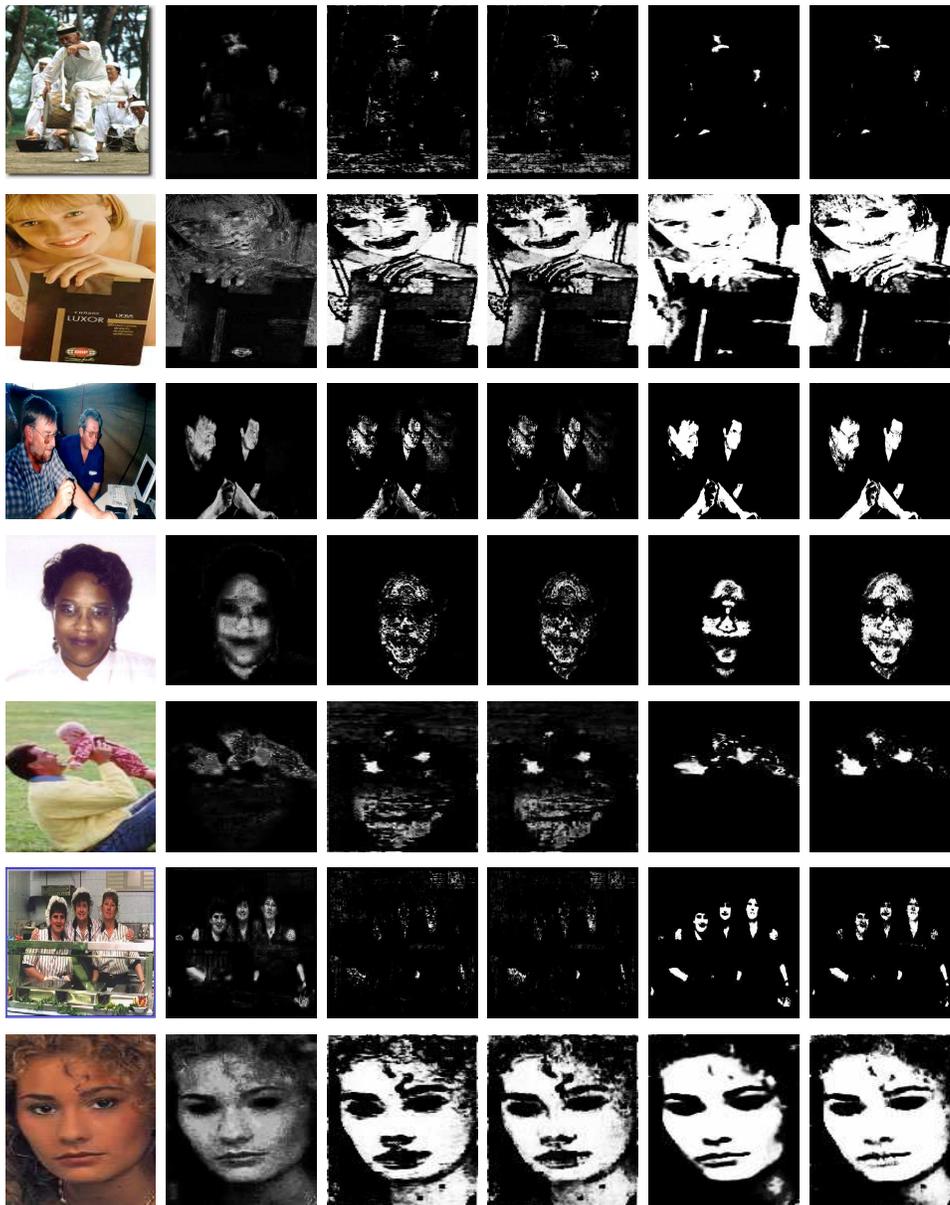


FIGURE 5.3 – Quelques entrées et sorties des différents modèles comparés. De gauche à droite, les sorties qui sont des images aux niveaux de gris correspondes aux modèles fondés sur le modèle de base, 2-BDTs, 1-BDT, diff-BDT et 1-BDT-diff-BDT respectivement.

2-BDTs et 1-BDT détectent, plus que les deux autres classifieurs basés sur un arbre, les pixels qui ont une couleur semblable à celle de la peau. Cela peut être expliqué par le fait que les critères d'optimisation à partir desquelles ces deux derniers arbres sont construits, n'exploitent pas la notion d'interclasse entre les distributions peau et non peau.

Le modèle de base comparé aux modèles basé sur 2-BDTs et 1-BDT, détecte en moins les pixels ayant une couleur semblable à celle de la peau. Par contre, la croyance en peau mesurée par les deux modèles précédent basés sur les arbres, est bien meilleure que celle donnée par le modèle de base lors d'une bonne détection d'un vrai pixel peau ou non peau.

Les résultats obtenus par les modèles basés sur diff-DBT et 1-BDT-diff-DBT, sont remarquablement performants et supérieurs à ceux des autres modèles. Ces performances peuvent toujours être expliquées par les critères d'optimisation utilisés. Dans le cas de ces deux classifieurs, les pixels sont mieux classés. Lorsqu'il s'agit d'un pixel non peau (respectivement non peau), le degré de croyance de la peau du pixel est proche de 1 (respectivement 0). Cela est nettement remarqué sur les images en sorties des classifieurs. Toutefois, nous notons que les résultats du détecteur basé sur 1-BDT-diff-DBT, sont les meilleurs.

Dans le but de réaliser une mesure quantitative des différents classifieurs étudiés, nous proposons de combiner les faux positifs et les vrais positifs obtenus au niveau de chaque détecteur de peau étudié. Ainsi pour les images aux niveaux de gris données par un détecteur, nous appliquons un seuil pour obtenir le masque (image binaire) de l'image en entrée. Les taux d'erreur pour tous les seuils sont récapitulés dans la courbe courbe ROC (Receiver Operating Characteristic) de chaque classifieur.

La Figure 5.4 représente les différentes courbes ROC des classifieurs étudiés basés sur un arbre ainsi que celle du classifieur fondé sur le modèle de base. Les axes des abscisses et des ordonnées représentent respectivement les proportions des faux positifs et des vrais positifs.

En comparant les différentes courbes ROC des modèles d'arbres par rapport à la courbe ROC du modèle de base, nous constatons que :

- Sur l'intervalle  $[0 \quad 0.065]$ , les courbes ROC des modèles 2-BDT et 1-BDT sont au-dessus de celle du modèle de base. Pour des valeurs supérieures à 0.065 de faux positifs,

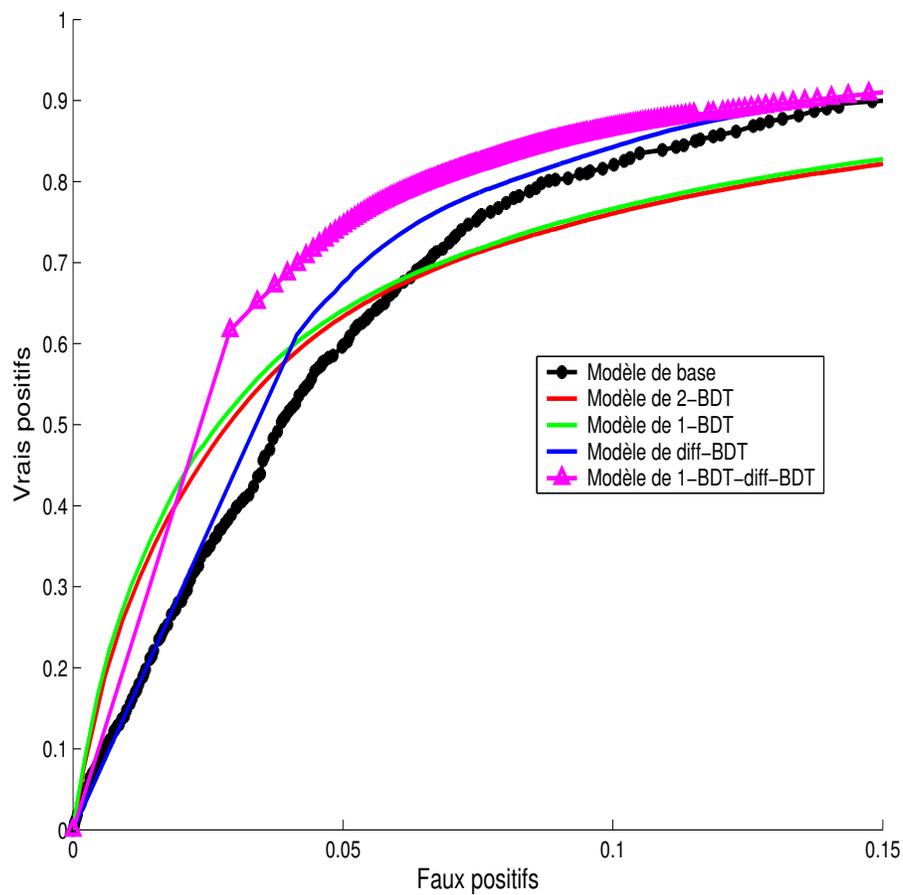


FIGURE 5.4 – Courbes ROC des classifieurs basés sur le modèle de base, 2-BDTs, 1-BDT, diff-BDT et 1-BDT-diff-BDT respectivement.

la courbe du modèle de base devient au dessus des modèles 2-BDT et 1-BDT. Ainsi la valeur du taux de faux positifs 6.5%, est une valeur critique qui tranche des performances des modèle 2-BDT, 1-BDT et de base. Cette valeur correspond à un taux de vrais positifs égal à 68% obtenu par ces différents modèles.

- Pour des valeurs de faux positifs faibles, les courbes des modèle diff-BDT et de base restent pratiquement identiques jusqu'à un faux positif égal à 0.025. A partir de cette dernière valeur, la courbe du modèle diff-BDT devient au dessus du celle du modèle de base.
- En ce qui concerne la courbe ROC du modèle 1-BDT-diff-BDT, elle est nettement au dessus de celle du modèle de base pour toutes les valeurs de faux positifs.

Concentrons nous maintenant sur les courbes ROC des modèles d'arbres, nous remarquons que :

- Quelque soit la valeur de faux positifs considérée, la courbe du modèle 1-BDT est toujours supérieure à celle du modèle 2-BDT. Cela montre qu'en comparant les deux modèles 2-BDT et 1-BDT, le deuxième modèle permet une meilleure détection de la peau par rapport au premier.
- Les courbes ROC des modèles 2-BDT et 1-BDT coupent respectivement la courbe ROC du modèle diff-BDT aux points (0.04, 0.57) et (0.04, 0.6). Pour des valeurs de faux positifs supérieures à 0.04, la courbe du modèle diff-BDT présente des performances remarquables vis-à-vis des modèles 2-BDT et 1-BDT.
- Les points (0.02, 0.4) et (0.02, 0.44) sont respectivement les points d'intersections entre les courbes des modèles 2-BDT et 1-BDT, et celle du modèle 1-BDT-diff-BDT. Pour toute valeur de faux positifs supérieure à 0.04, le modèle 1-BDT-diff-BDT présente des taux de vrais positifs nettement supérieurs par rapport aux modèles 2-BDT et 1-BDT.
- Pour toute valeur de faux positifs, le modèle 1-BDT-diff-BDT offre de bonnes détections (vrais positifs) par rapport au modèle diff-BDT.

En conclusion, nous déduisons que les performances offertes par les différents modèles d'arbre à bi-probabilité peau et non peau proposés dans ce mémoire, sont considérablement supérieures au modèle d'arbre noté 2-BDT. En particulier, partant du modèle 1-BDT, passant par le modèle diff-BDT et arrivant au modèle 1-BDT-diff-BDT, la qualité de la détection de la peau s'améliore d'un modèle à un autre. Le détecteur de peau fondé sur le modèle 1-BDT-diff-BDT présente les meilleures performances en discriminant au mieux

les pixels peau de ceux non peau ; pour un taux de faux positifs égal à 12.5%, un taux de vrais positifs égal à 90% est atteint par ce modèle. De plus le modèle 1-BDT-diff-BDT (arbre de dépendance à bi-probabilité peau et non peau Discriminatoire) présente une détection de peau remarquablement supérieure à celle donnée par le modèle de base. Ce dernier offre un taux de détection de vrais positifs égal à 90% pour un taux de faux négatifs toléré égal à 14.2%.

### 5.3 Deuxième étude comparative

Dans cette étude comparative, des classifieurs probabilistes de peau/non peau basés sur les arbres de dépendances des sont évalués. Plus précisément, nous considérons ceux issus des mélanges des arbres de dépendances optimaux basés sur le critère d'optimisation proposé par Chow Liu. Ici, nous étudions quatre mélanges possibles que nous notons :

- Mix-1-ArbreROpt : dans le cas où un seul arbre est considéré ;
- Mix-2-ArbreROpts : dans le cas où deux arbres de dépendances optimaux sont mélangés ;
- Mix-3-ArbreROpts : dans le cas où trois arbres de dépendances optimaux sont mélangés ;
- Mix-4-ArbreROpts : dans le cas où quatre arbres de dépendances optimaux sont mélangés.

Afin d'enlever toute ambiguïté, nous soulignons que lorsque nous disons qu'un classifieur est basé sur un arbre de type arbre de Chow Liu, nous entendons par cela que deux arbres de type Chow Liu sont utilisés. L'un des deux arbres indexe la probabilité peau alors que l'autre arbre indexe la probabilité non peau.

Par un apprentissage automatique, nous avons pu définir les quatre arbres de dépendances optimaux qui constitueront les bases de nos classifieurs de peau/non peau. A noter que les procédures de constructions de ces différents classifieurs ont été décrites dans le chapitre 4 section 4.3 pour un nombre  $K$  quelconque d'arbres mélangés ( $K \in N$ ). Aussi, l'erreur de tolérance utilisée pour la définition des différents arbres sera résumée dans le Tableau 5.2. Les Figures 5.5 et 5.6 représentent les quatre arbres de dépendances optimaux peau (respectivement non peau) utilisés par les différents classifieurs. Les arbres représentés sont ordonnés selon leurs poids de 1 à 4. Ils vérifient :

$$W^{T^{i1}} \geq W^{T^{i2}} \geq W^{T^{i3}} \geq W^{T^{i4}} \quad , \quad i \in \{1, 2\} \quad (5.2)$$

où la valeur entière  $i$  correspond au label des classe peau ('1') et non peau ('2'), Alors que  $T^{ij}$  est un arbre de dépendances optimal pour tout  $j = 1, 2, 3, 4$ .

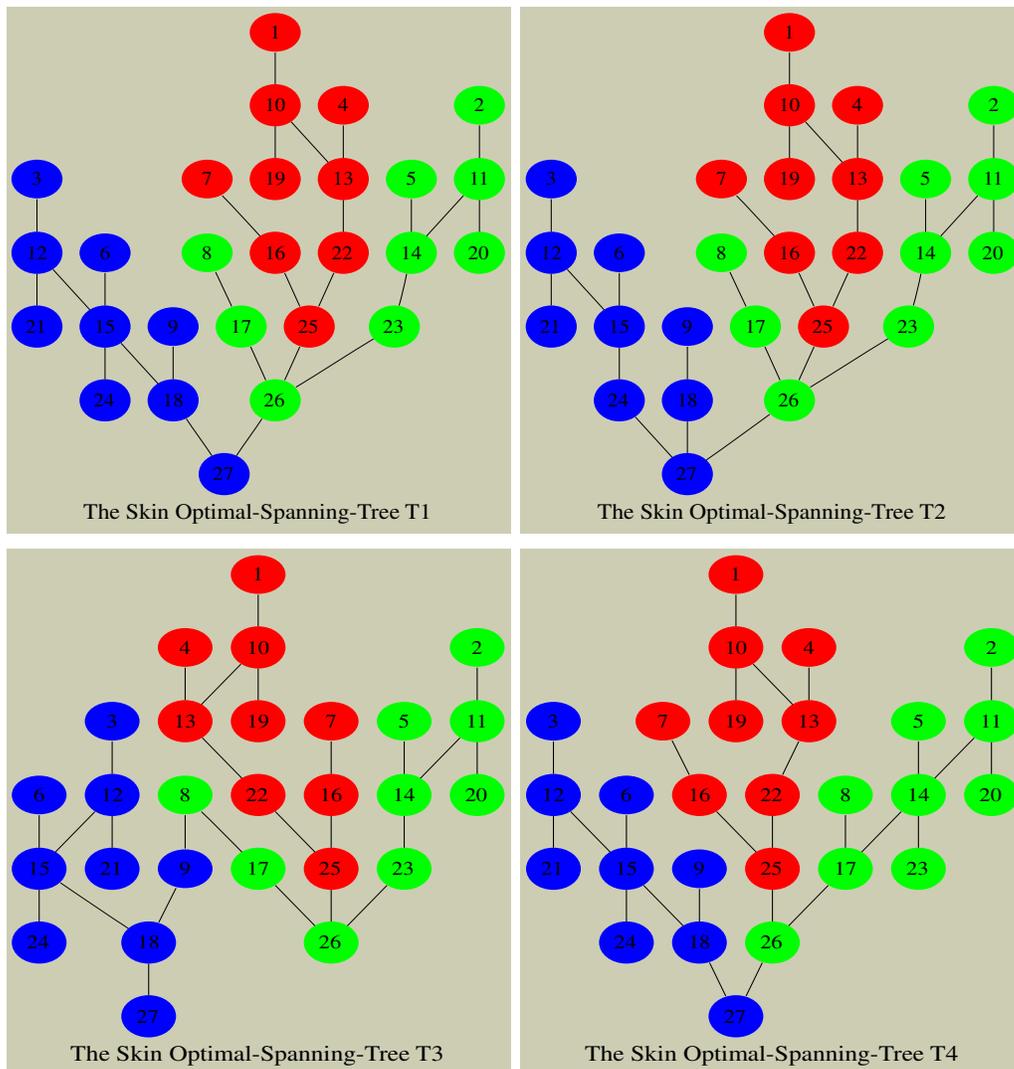


FIGURE 5.5 – Les quatre arbres de dépendances optimaux indexant la classe peau, obtenus par l'apprentissage.

Nous attirons l'attention que les arbres utilisés ne sont pas de type "arbres colorés". L'usage de la couleur a pour unique intention la mise en valeur les dépendances retenues sur un arbre.

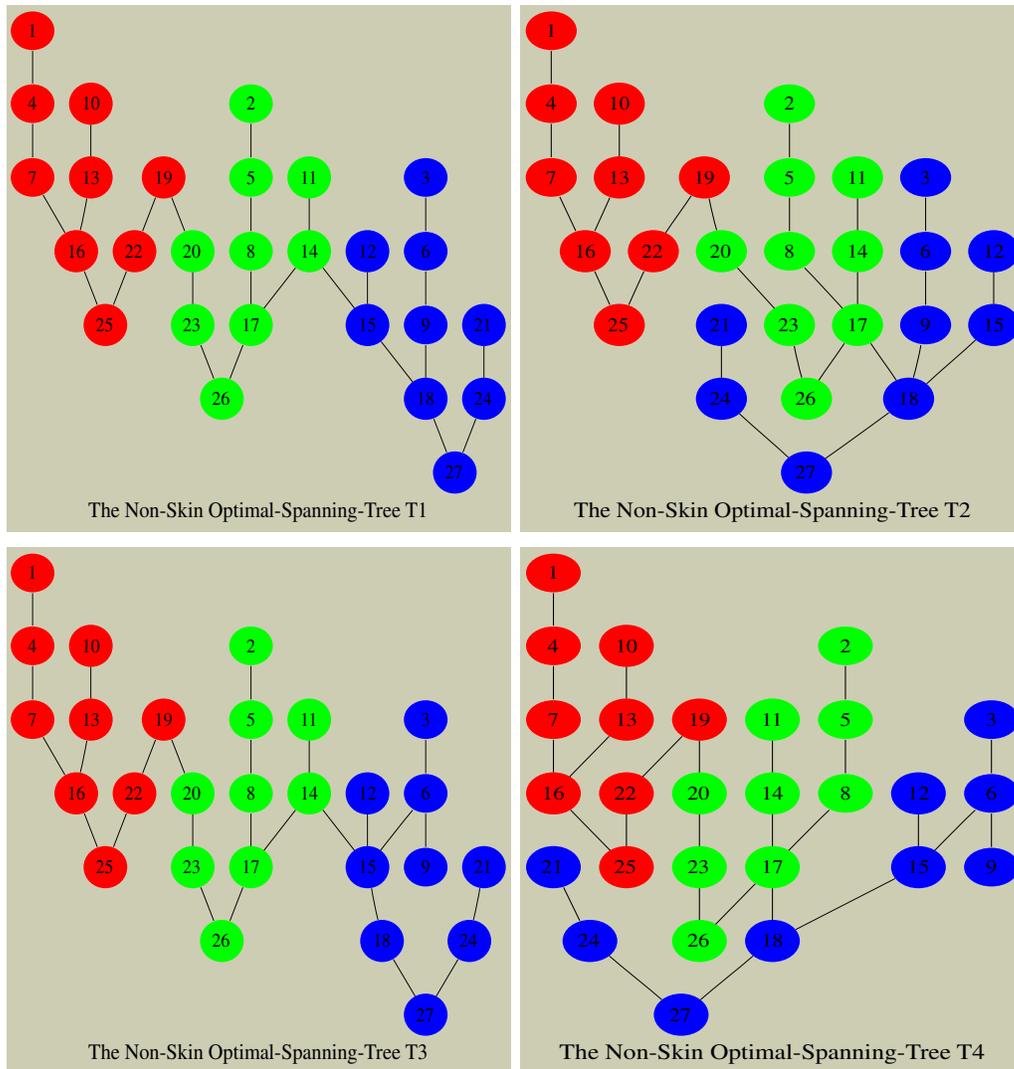


FIGURE 5.6 – Les quatre arbres de dépendances optimaux indexant la classe non peau, obtenus par l'apprentissage.

Les Figures 5.5 et 5.6 prouvent que la décomposition à laquelle nous avons procédé, concernant le vecteur des observations, est justifiée. En effet, les noeuds rouges, bleus et verts ne sont reliés entre eux qu'en fonction d'une arête pour assurer une structure connexe de l'arbre de recouvrement. Par contre, les noeuds de la même couleur, sont tous reliés entre eux. Cela est aussi en adéquation avec le fait que quelques travaux considèrent

chaque plan de couleur à part pour effectuer leurs traitements. L'explication d'un tel procédé est donnée par la forte corrélation qui existe entre les mêmes composantes d'une même couleur rouge, bleue ou verte.

Procédons aux mélanges des différents arbres peau (resp. non peau) sélectionnés. A cette fin, la table 5.1 définit les quatre ensemble d'arbres  $\Theta^K$  ( $K = 1, 2, 3, 4$ ) utilisés à chaque mélange.

TABLE 5.1 – Ensemble d'arbres considérés par chaque mélange étudié.

<b>Modèle de mélange</b>	<b>Ensemble d'arbres mélangés</b>
Mix-1-ArbreROpt	$\Theta^1 = \{T^1\}$
Mix-2-ArbreROpts	$\Theta^2 = \{T^1, T^2\}$
Mix-3-ArbreROpts	$\Theta^3 = \{T^1, T^2, T^3\}$
Mix-4-ArbreROpts	$\Theta^4 = \{T^1, T^2, T^3, T^4\}$

L'erreur de tolérance est utilisée dans la sélection des quatre arbres de dépendances optimaux. La table 5.2 résume l'erreur de tolérance considérée par chaque ensemble d'arbres à mélanger.

TABLE 5.2 – Erreurs de tolérance considérées par les différents mélanges.

<b>Modèle de mélange</b>	<b>Erreur de tolérance</b>
Mix-1-ArbreROpt	0
Mix-2-ArbreROpts	0
Mix-3-ArbreROpts	0.00004130
Mix-4-ArbreROpts	0.00004830

L'arbre obtenu par le mélange Mix-1-ArbreROpt n'est autre que celui représenté dans les Figures 5.5 et 5.6 par l'arbre de dépendances  $T1$ . Quant aux arbres de mélanges obtenus par chaque mélange Mix-K-ArbreROpts considéré ( $K = 2, 3, 4$ ), ils sont représentés à la Figure 5.7, et ce dans les cas des deux classes peau et non peau. Sur cette figure encore, nous remarquons que les noeuds appartenant à la même composante de couleur

sont fortement reliés entre eux. La relation entre deux noeuds de couleurs différentes, sert comme une jonction pour garantir une connexité de la structure de l'arbre recouvrant du mélange. Il est à noter que ces différents arbres permettent de mettre en évidence les dépendances les plus importantes sur le graphe de mélange, résultant du mélange conjoint des structures et des probabilités des différents arbres considérés.

Dans un premier temps, nous considérons des classifieurs où chacun d'entre eux est fondé sur un couple d'arbres. Chaque couple d'arbres est constitué de l'arbre de dépendances optimal peau  $T^{1j}$  et l'arbre de dépendances optimal non peau  $T^{2j}$ ; avec  $j = 1, 2, 3, 4$  désigne le numéro de l'arbre utilisé.

Quelques résultats obtenus par ces différents classifieurs sont représentés dans la Figure 5.8. Les entrées des classifieurs sont des images couleurs données par la première colonne de la figure. Quant aux sorties des classifieurs, elles sont représentées par les images aux niveaux de gris. De gauche à droite, les images aux niveaux de gris correspondent aux classifieurs basés sur les couples d'arbres  $(T^{1j}, T^{2j})$ ; avec  $j$  égal à 1, 2, 3 et 4 successivement. Nous constatons que chacun des différents classifieurs considérés présente une amélioration vis-à-vis des autres classifieurs. C'est d'ailleurs dans l'intention de bénéficier des apports de chacun des arbres de dépendances optimaux possibles et d'exploiter les connaissances sur chacun d'eux, que l'idée du mélange est née.

Considérons maintenant les classifieurs issus des mélanges Mix-K-ArbreROpts ( $K = 2, 3, 4$ ). Quelques entrées et sorties de ces classifieurs sont données par la Figure 5.9. La première colonne de cette figure représente les images couleurs placées en entrées des différents classifieurs comparés. Les images en sorties des détecteurs basés sur les Mix-K-ArbreROpts, sont représentées sur les colonnes deux ( $K = 2$ ), trois ( $K = 3$ ) et quatre ( $K = 4$ ).

Les qualités des différentes sorties des classifieurs basés sur les mélanges, met en valeur les performances offertes par de tels classifieurs. En particulier, une qualité de classification remarquable est soulignée par le classifieur fondé sur le Mix-3-ArbreROpts.

Visuellement, le choix du mélange d'arbres de dépendances optimaux est justifié par une simple comparaison entre les images aux niveaux de gris des Figures 5.8 et 5.9. Plus précisément, nous optons pour le mélange de trois arbres de dépendances optimaux.

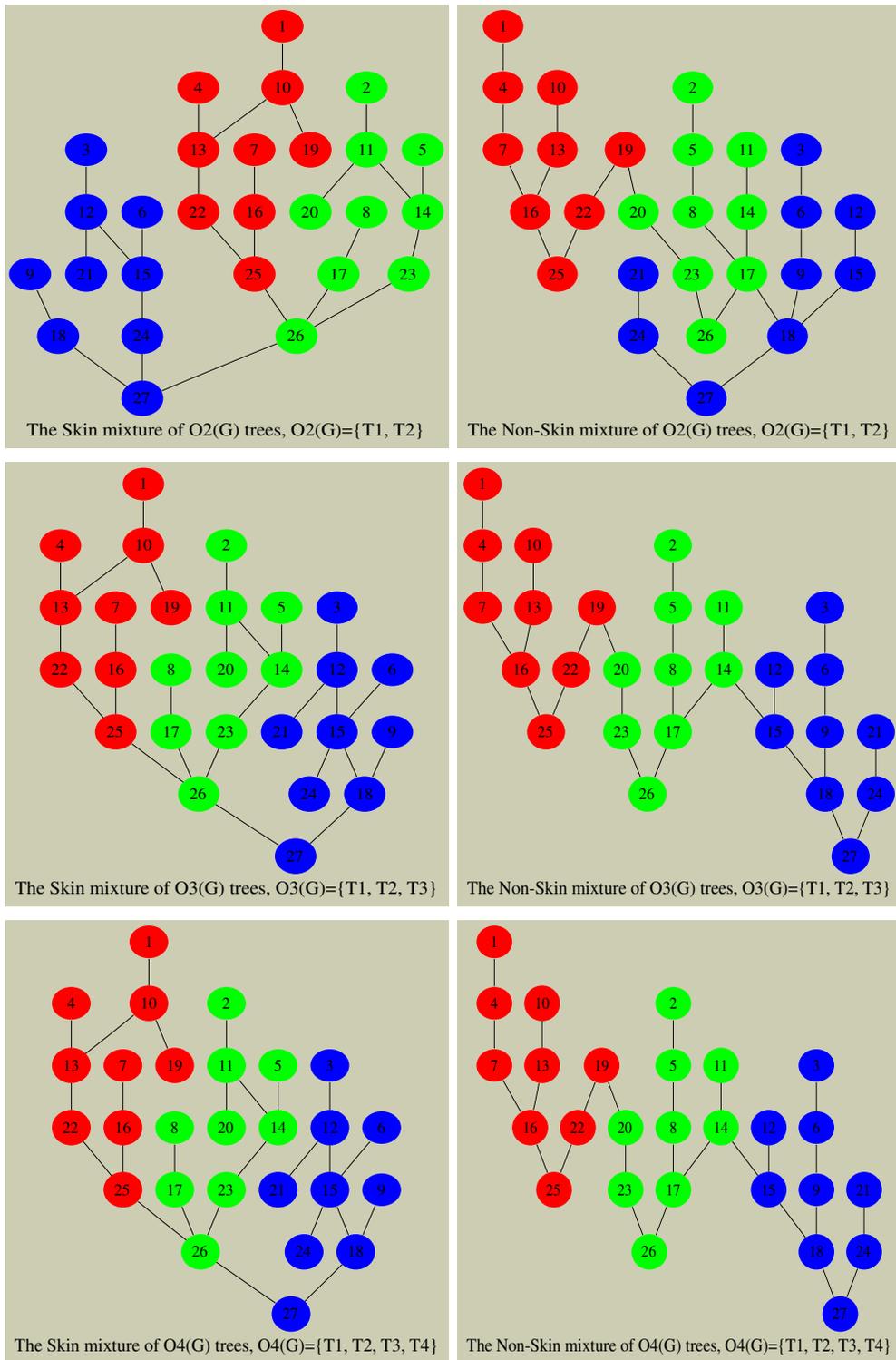


FIGURE 5.7 – Les arbres de mélanges indexant la distribution de peau ou celle de non peau, obtenus par les trois mélanges considérés ( $K = 2, 3, 4$ ).

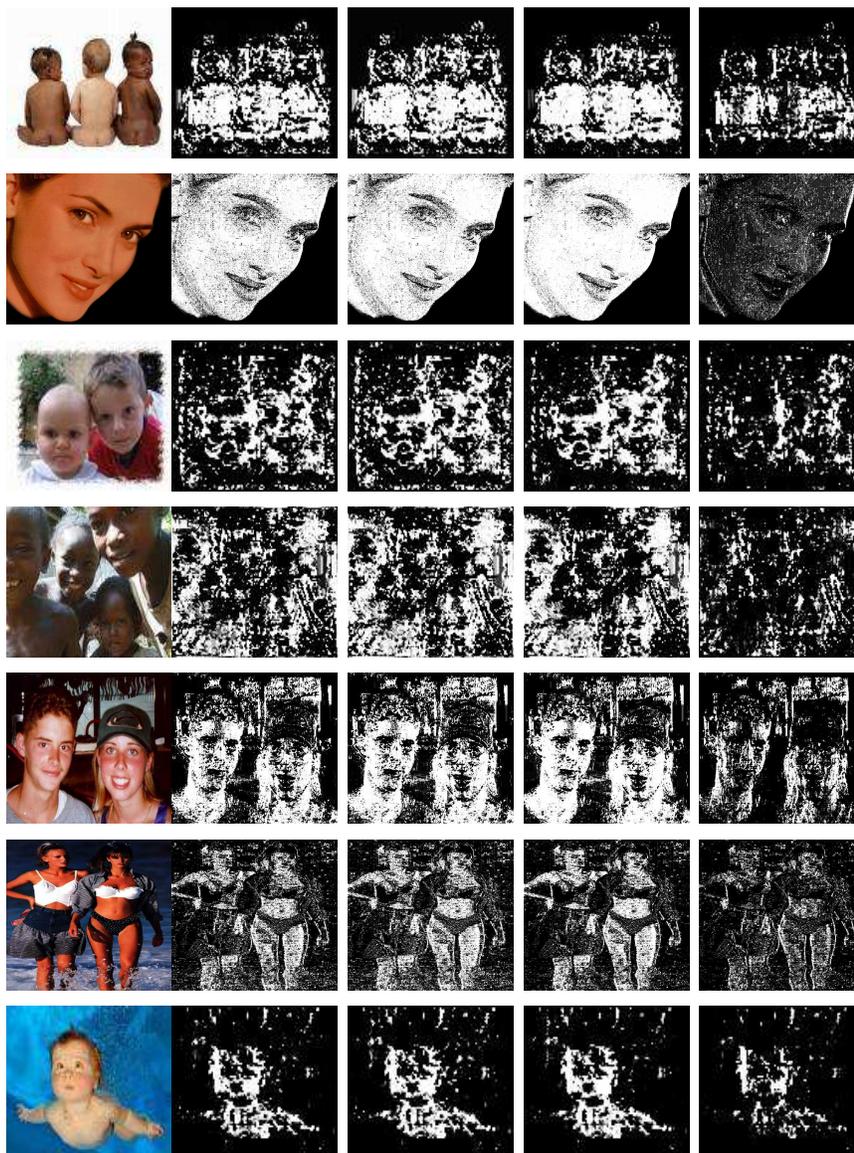


FIGURE 5.8 – Quelques entrées et sorties des classifieurs dont chacun est basé sur un couple d'arbres de peau et non peau. Première colonne représente les images couleurs (entrées). Ensuite les images aux niveaux de gris obtenues par les différents classifieurs sont données. Les sorties des classifieurs sont ordonnées de gauche à droite selon l'ordre croissant des numéros des arbres considérés  $T^K$  ( $K = 1, 2, 3, 4$ ).

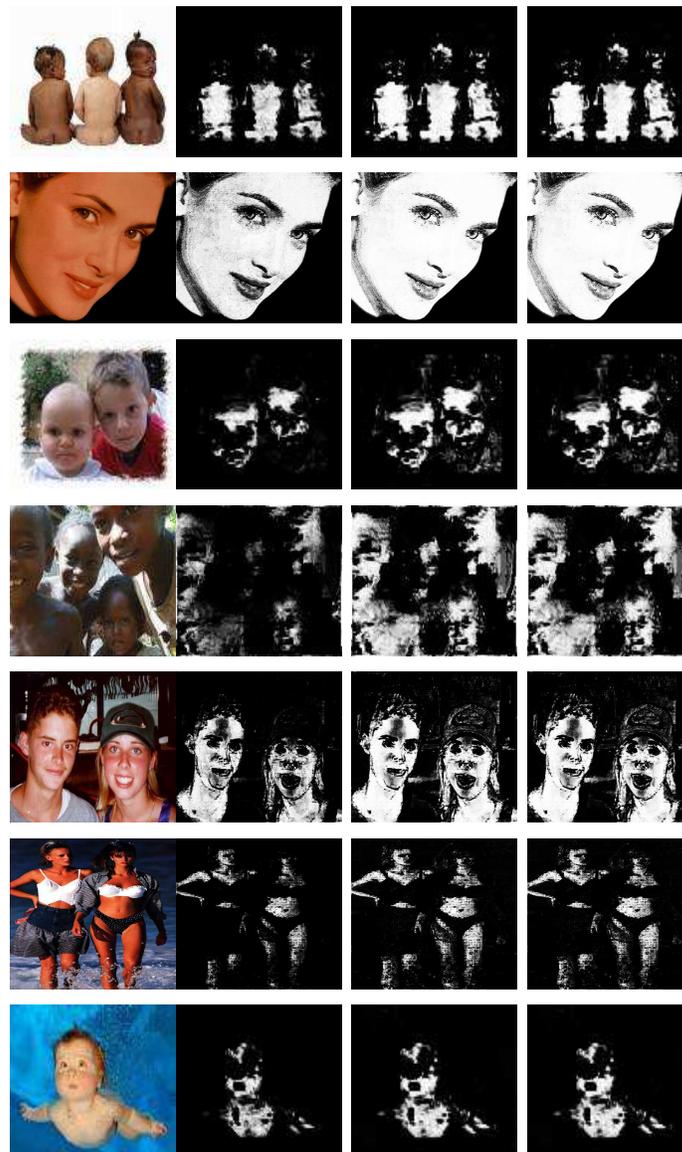


FIGURE 5.9 – Quelques entrées et sorties des classifieurs basés sur les mélanges d’arbres. La première colonne représente les images couleurs correspondant aux entrées des classifieurs. Les sorties des classifieurs sont ordonnées de gauche à droite selon le nombre d’arbres mélangés  $\Theta^K$  ( $K = 2, 3, 4$ ).

Une évaluation quantitative des différents classifieurs étudiés dans la présente section est schématisée sur la Figure 5.10.

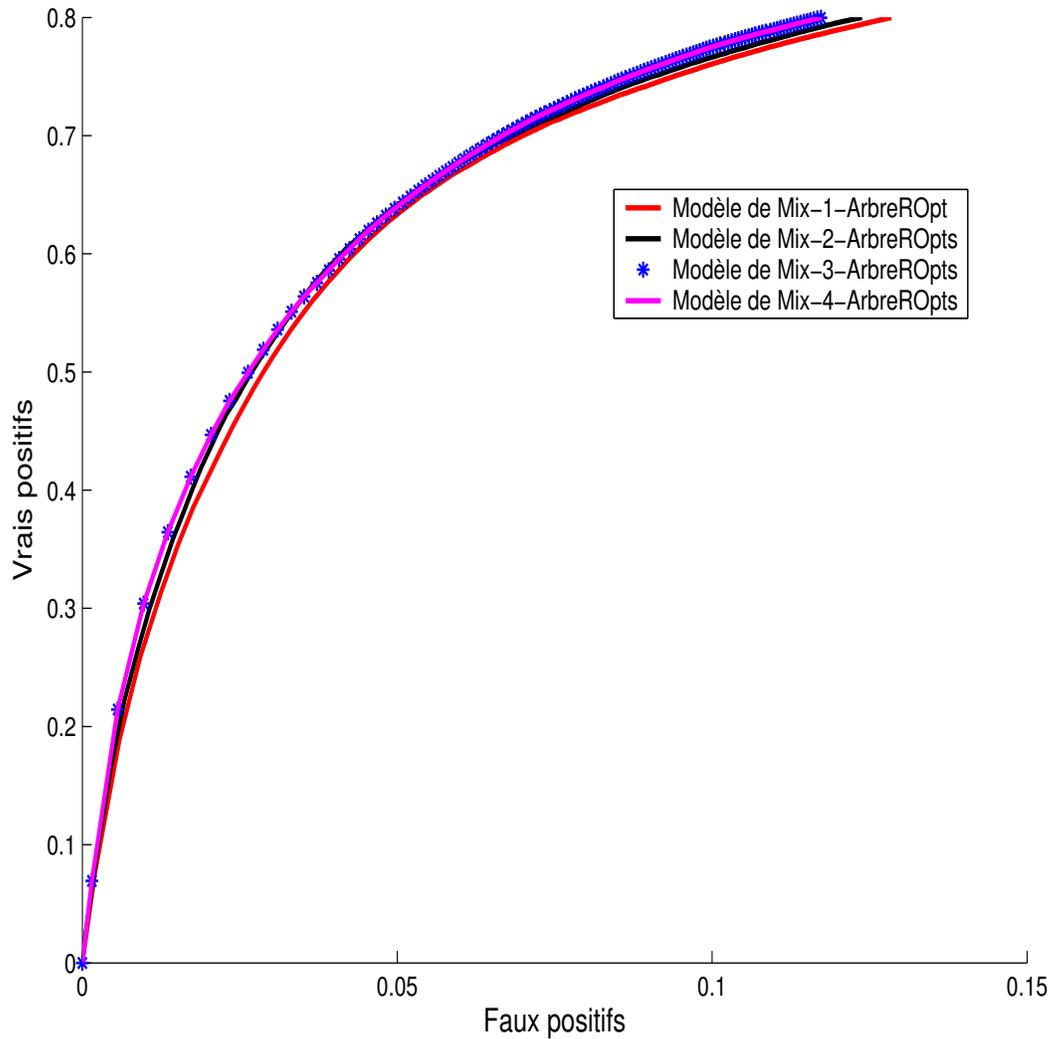


FIGURE 5.10 – Courbes ROC des classifieurs étudiés fondés sur les modèles de mélanges d’arbres.

La première des remarques à noter sur les courbes ROC de la Figure 5.10, est que les courbes des modèles de mélanges Mix- $K$ -ArbreROpts sont toujours au-dessus de la courbe du mélange Mix-1-ArbreROpt (courbe en rouge), et cela pour tout  $K = 2, 3$  et 4. Si nous avons monté dans le chapitre précédent, que le modèle de mélange proposé dans le cas des arbres de type Chow Liu améliorera théoriquement la qualité de la détection de la peau, les courbes ROC justifient que les résultats théoriques et expérimentales concordent.

Aussi nous remarquons que les performances des mélanges données par leurs courbes augmentent parallèlement avec le nombre d'arbres à mélanger. Cela est justifié par le fait que chaque arbre ajouté à un nouveau mélange considéré emmagasine des informations pertinentes qui ont été ignorées par le mélange lui précédant.

Quelques statiques sur les courbes ROC des modèles de mélanges. Pour un taux de faux positifs égal à 10%, les taux de vrais positifs obtenus par les modèles des mélanges Mix- $K$ -ArbreROpts, pour  $K = 1, 2, 3$ , et 4, sont respectivement 75.5%, 76.5%, 77.8% et 77.8%.

Les courbes ROC des modèles Mix-3-ArbreROpts et Mix-4-ArbreROpts sont confondues. Cela peut être expliqué par le fait qu'en dépit de considérer une valeur d'erreur de tolérance supérieure à un certains seuil, les arbres à ajouter au mélange ne contiennent par des informations supplémentaires qui caractérisent le système de détection de la peau étudié. Ainsi sur le corpus des images utilisé, il s'avère que le mélange de trois arbres de type de Chow Liu est suffisant pour caractériser un bon détecteur de peau dans les images couleurs.

## 5.4 Troisième étude comparative

Dans cette troisième étude comparative des classifieurs probabilistes basés sur les arbres, nous nous intéressons en particulier à ceux utilisant la combinaison d'arbres telle que présentée au chapitre 4 section 4.4.2. Plus particulièrement, nous comparons des classifieurs basés sur les Meilleurs Arbres Recouvrants proposés dans ce mémoire (chapitre 3 section 3.4). Ici, nous étudions deux mélanges possibles que nous notons :

- Comb-1-ArbreROpt : dans le cas où un seul arbre est considéré ;
- Cpob-3-ArbreROpts : dans le cas où trois Meilleurs Arbres Recouvrants possibles sont mélangés.

Quelques résultats obtenus par les deux modèles étudiés sont illustrés par la Figure 5.11. L'entrée de chaque modèle est une image couleur et la sortie est une image au niveau de gris. La deuxième colonne représente les sorties du modèle Comb-1-ArbreROpt alors que la troisième colonne donne les sorties du modèle Comb-3-ArbreROpts.

La qualité des images en sorties des modèles de combinaisons, montre que la détection de la peau est mieux accomplie par le modèle Comb-3-ArbreROpts. Quelques pixels peau ou non peau mal classés par le modèle Comb-1-ArbreROpt sont bien classés par l'autre modèle.

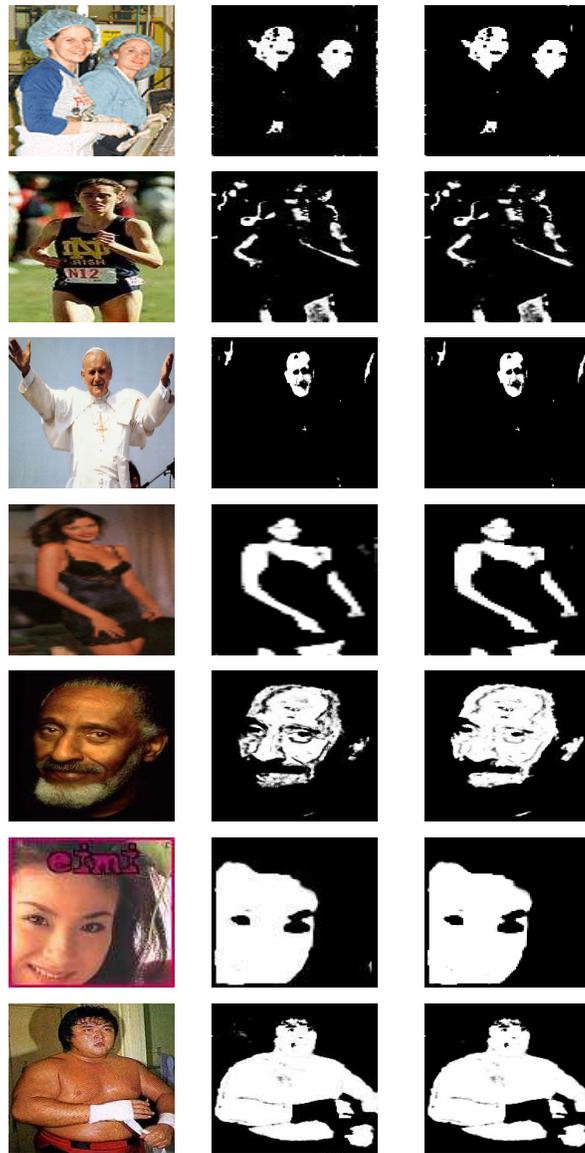


FIGURE 5.11 – Quelques entrées et sorties des classifieurs de peau fondé sur les combinaisons d’arbres. La première colonne représente les images couleurs correspondant aux entrées des classifieurs. La deuxième et la troisième colonne correspondent respectivement aux sorties des classifieurs fondés sur Comb-1-ArbreROpt et Comb-3-ArbreROpt.

Ce constat peut être expliqué par le fait que des informations pertinentes en défaut dans le modèle Comb-1-ArbreROpt ont été ajoutées au modèle Comb-3-ArbreROpts. Ces informations sont emmagasinées dans les deux autres Meilleurs Arbres Recouvrants possibles qui ont été considérés par l'approche Comb-3-ArbreROpts.

La Figure 5.12 représente une évaluation quantitative d'un classifieur basé sur un Meilleur Arbre Recouvrant et d'un autre basé sur la combinaison de trois Meilleurs Arbres Recouvrants.

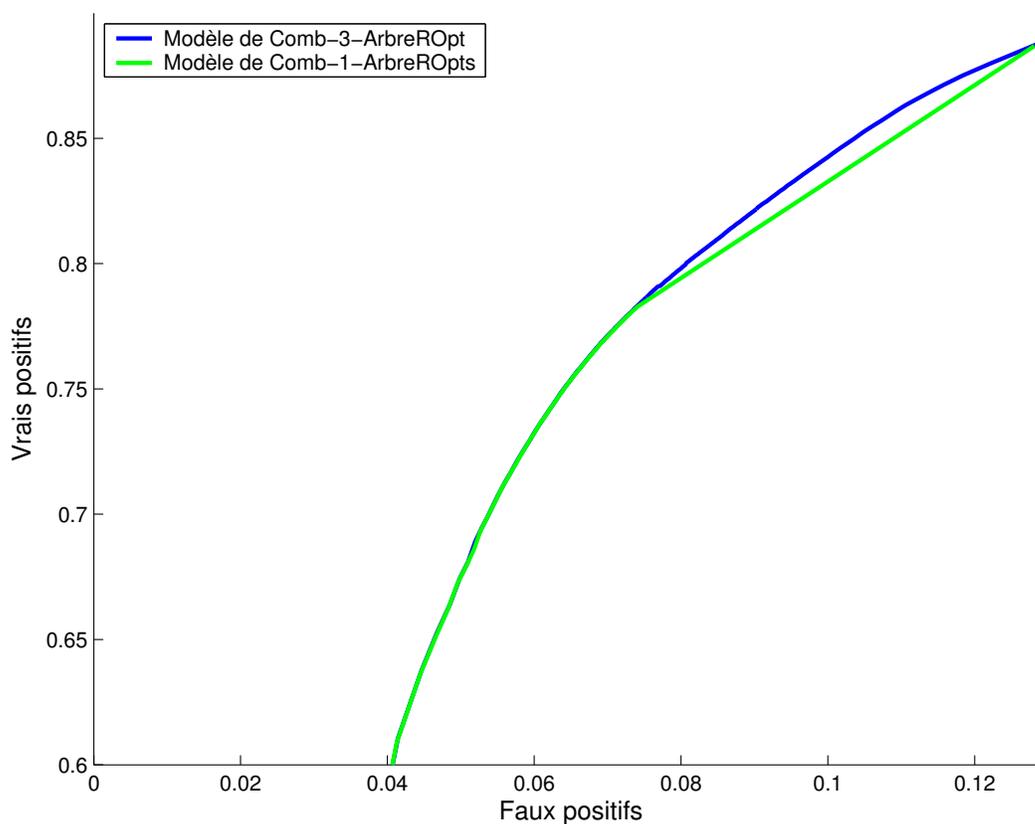


FIGURE 5.12 – Courbes ROC des classifieurs étudiés fondés sur les combinaisons des Meilleurs Arbres Recouvrants.

Dans notre cas, les courbes ROC des deux modèles comparés sont confondues pour toute valeur de faux positifs hors l'intervalle  $[0.075; 0.12]$ . Sur ce dernier intervalle, la courbe ROC du modèle Comb-3-ArbreROpts est au-dessus de celle du modèle Comb-1-ArbreROpt.

Cela signifie que les résultats obtenus par le modèle de combinaison des trois arbres sont supérieurs à ceux obtenus par le modèle de combinaison d'un seul arbre. Pour des taux de faux positifs égaux à 8%, 10.5%, 12%, les taux de vrais positifs pour chaque couple de modèles (Comb-1-ArbreROpt, Comb-3-ArbreROpts) sont respectivement (79.2%, 80%), (84.1%, 85%) et (87%, 87%).

Nous concluons que l'approche de combinaison proposé au chapitre 4 section 4.4.2, permet d'améliorer la qualité de la détection des pixels peau et ceux non peau, et cela vis-à-vis du modèle où un seul arbre est considéré.

## 5.5 Discussion

Le long du présent chapitre, nous avons essayé de comparer les performances des différents modèles de la peau présentés dans ce mémoire. Les approches proposées sont de trois types : (1) méthodes fondées sur un arbre indexant conjointement les distributions des probabilités peau et non peau, (2) approche du mélange des arbres et (3) celle des combinaisons des arbres. Dans ce contexte, nous avons réalisé trois études comparatives dont chacune concerne un entité des trois types proposés.

Afin de pouvoir comparer tous les résultats obtenus par les différentes études menées, nous avons adopté un protocole expérimental commun entre ces différentes études. La plus large base des images dédiée à l'évaluation des modèles de la détection de la peau, a été utilisé ici comme corpus des images pour l'apprentissage et le test des approches évaluées. Ce corpus englobe des images contenant de la peau et d'autres sans peau prises dans différentes conditions de prises de vues et comportent des personnes ayant des couleurs de peau variables. En outre, l'espace de couleur utilisé par les différentes méthodes est RVB. Par ailleurs, deux méthodes sont ajoutées à la liste de celles étudiées afin de comparer les performances de nos modèles par rapport à celles données par l'existant. Il s'agit de l'approche donnée dans [86], que nous notons modèle de base, et le modèle de l'arbre de Chow Liu défini à [92].

Le tableau 5.3 résume les résultats obtenus par les différentes approches comparées dans les trois études comparatives menées, dans le présent chapitre.

Les statistiques données par le tableau 5.3, prouvent expérimentalement que les approches proposées dans cette thèse améliorent la qualité de la détection de la peau dans

TABLE 5.3 – Tableau récapitulatif des résultats obtenus par les trois études comparatives menées.

Méthode fondée sur	Vrais positifs (VP)	Faux positifs (FP)
<u>Arbre unique et modèle de base</u>		
2-BDTs	75.5%	10%
1-BDT	76%	10%
diff-BDT	84.1%	10.5%
1-BDT-diff-BDT	90%	12.5%
Modèle de base [86]	90%	14.2%
<u>Mélange d'arbres optimaux possibles de type Chow Liu</u>		
Mix-2-ArbreROpts	76.5%	10%
Mix-3-ArbreROpts	77.8%	10%
Mix-4-ArbreROpts	77.8%	10%
<u>Combinaison de Meilleurs Arbres Recouvrants possibles</u>		
Comb-3-ArbreROpts	85%	10.5%

les images couleurs, et cela en comparaison avec les modèles de base existants. En effet, les performances des modèles comparés, ont montré en termes de qualité et de quantité des résultats mesurés que :

- Les modèles d'arbres de dépendances à bi-probabilité peau et non peau offrent de meilleures statistiques vis-à-vis du modèle d'arbre de Chow Liu. Cela est justifié par le fait que l'arbre indexant conjointement les probabilités de peau et non peau, exploite les interactions entre les classes peau et non peau. La gestion des inters et intras classe entre les deux classes citées, favorise une meilleure modélisation de la distribution de probabilité pour chacune des deux classes.
  - Le modèle du mélange des arbres optimaux vérifiant le critère d'optimisation donné par Chow Liu, permet d'offrir de meilleurs résultats par au modèle où un seul arbre Chow Liu est considéré. En particulier l'utilisation de l'erreur de tolérance a permis de prendre en considération de nouvelles arbres de dépendances optimaux qui emmagasine des propriétés pertinentes sur les distributions peau et non peau.
  - Le modèle de combinaison des Meilleurs Arbres Recouvrants améliore d'avantage la qualité de détection de la peau donnée par un seul Meilleur Arbre Recouvrant considéré.
- L'approche proposé, dit modèle de l'arbre de dépendances à bi-probabilité peau et non peau Discriminatoire offre une meilleure discrimination des classes peau et non peau, avec un taux de détection égal à 90% lorsqu'un taux de faux positifs égal à 12.5% est toléré. Par ailleurs, une combinaison des tous les arbres de dépendances à bi-probabilité peau et non peau Discriminatoires possibles sur un graphe, offrira une qualité de détection de la peau supérieure.



---

## CONCLUSION GÉNÉRALE

La présente thèse est consacrée à la résolution de l'une des célèbres problématiques de traitement des images. Il s'agit de la classification des pixels des images en peau et non peau. Cette tâche facile à réaliser par l'oeil humain est restée une problématique ardue à résoudre automatiquement par un ordinateur. La diversité des conditions de prises de vues des images, celle de la présence de plusieurs couleurs de la peau humaine ainsi que des objets ayant une couleur semblable à celle de la peau, tous ces facteurs ont rendu la détection de la peau un sujet non sans difficulté. Nous avons essayé, à travers la présente thèse, de définir et valider un modèle probabiliste robuste pour la détection de la peau dans les images couleurs.

La détection de la peau dans les images couleurs gagne une place remarquable dans le domaine de la recherche scientifique. Cela est principalement dû à son rôle primordial comme une phase de prétraitement importante dans plusieurs applications [4][5][6][7][8][9][10][11]. Pour pouvoir déterminer un modèle robuste de la détection de la peau, nous avons présenté dans le premier chapitre un ensemble de méthodes paramétriques et non paramétriques existantes. Une grande masse des approches de la détection de la peau sont fondées sur l'indice couleur comme l'information discriminatoire des pixels peau et des pixels non peau dans une image. Dans la littérature, le modèle de peau basé sur la couleur, qui a présenté de meilleures performances en termes de bonne détection est celui présenté dans [86]. Nous avons adopté ainsi cette caractéristique, la couleur, pour répondre à notre objectif dans le cadre d'une classification supervisée. De plus, nous avons adopté des modèles pro-

babillistes indexés par les arbres pour définir un modèle de peau. L'idée cruciale derrière le choix de tels modèles, se résume dans leur efficacité, simplicité et intuitivité pour traiter les données graphiques (images).

Nous avons pu tirer profit des modèles des arbres existants indexant des probabilités des classes peau ou non peau, pour déterminer une nouvelle génération d'arbres que nous avons nommée " arbre de dépendances à bi-probabilité peau et non peau ". Il s'agit d'un arbre qui permet d'indexer conjointement les distributions de probabilités peau et non peau. L'intérêt de ce modèle d'arbre réside dans l'exploitation des connaissances et interactions entre les deux distributions de probabilités en question, et ceci pour modéliser au mieux ces deux distributions de probabilités. Nous avons pu définir trois différents modèles de peau fondés sur l'arbre de dépendances à bi-probabilité peau et non peau, à savoir les modèles dits Simple, Meilleur Arbre Recouvrant et Discriminatoire. Le concept de l'arbre de recouvrement optimal qui trouve son application dans plusieurs secteurs d'activité, a été utilisé dans les définitions de nos modèles. Chaque modèle proposé se distingue par un critère d'optimisation que nous avons déterminé pour décrire les caractéristiques des deux distributions peau et non peau.

Nous avons constaté que l'unicité des arbres optimaux sur un graphe n'est pas toujours garantie. De ce principe, il fallait exploiter toutes les informations apportées par les différents arbres possibles. C'est dans cette perspective que nous avons défini des modèles de mélanges ou de combinaisons d'arbres de dépendances recouvrants. Des études théoriques, menées et présentées au chapitre 4, nous ont permis de définir des modélisations originales de peau théoriquement robustes. Il est à noter que le mélange proposé se différencie du mélange classique par le mélange conjoint des structures des arbres de dépendances optimaux possibles et des probabilités portées par ces arbres, et ceci dans l'intention de définir un arbre optimal idéal indexant les distributions peau et non peau. Quant à l'approche de combinaison proposée dans ce mémoire, c'est une instance des modèles de combinaisons parallèles. Ce type de combinaisons permet de traiter les différentes sorties des classifieurs combinés, à pied d'égalité lorsqu'aucune pré-connaissance sur le comportement des classifieurs n'est disponible.

La validation et l'évaluation de nos différents modèles fondés sur, un seul arbre, des mélanges d'arbres ou encore des combinaisons d'arbres pour la détection de la peau dans les images couleurs, ont été réalisées sur le corpus Compaq. Ce dernier contient des images

remplissant des différents facteurs qui rendent la tâche de la détection de la peau difficile à résoudre. Les différentes études expérimentales comparatives des modèles proposés dans cette thèse, ont montré la supériorité de nos différents modèles vis-à-vis de celui donné dans [86] et celui fondé sur les arbres de probabilité donné dans [92].

Le tableau 5.4, ci-après, synthétise une comparaison entre nos différents modèles fondés sur les arbres pour indexer les probabilités peau et non peau.

TABLE 5.4 – Comparaison entre les modèles de détection de la peau proposés dans cette thèse.

Méthode fondée sur	Vrais positifs (VP)	Faux positifs (FP)
<u>Un arbre de dépendance à bi-probabilité peau et non peau</u>		
– Simple	76%	10%
– Meilleur Arbre Recouvrant	84.1%	10.5%
– Discriminatoire	90%	12.5%
<u>Mélange d'arbres optimaux possibles de type Chow Liu</u>		
– Mélange de deux arbres	76.5%	10%
– Mélange de trois arbres	77.8%	10%
– Mélange de quatre arbres	77.8%	10%
<u>Combinaison de Meilleurs Arbres Recouvrants possibles</u>		
– Combinaison de trois arbres	85%	10.5%

A partir de ce tableau, nous constatons que les modèles de mélanges (respectivement les modèles de combinaisons) permettent d'améliorer la qualité de la détection de la peau vis-à-vis du modèle où un unique arbre est considéré. Ainsi, nous constatons qu'aussi bien pour le mélange que la combinaison des arbres, les études théoriques et expérimentales concordent.

A partir des résultats obtenus, l'approche de l'arbre de dépendances à bi-probabilité peau et non peau Discriminatoire est considéré la plus robuste pour séparer les pixels peau de ceux non peau d'une image couleur. La performance de ce dernier modèle est justifiée par le critère d'optimisation utilisé pour la sélection de cet arbre. Ce critère formalise

explicitement les interactions d'inters et intras classe entre les classes peau et non peau. Par ailleurs, le modèle de combinaison de tous les dépendances à bi-probabilité peau et non peau Discriminatoires possibles sur un graphe, permettra d'améliorer d'avantage la détection des pixels peau et non peau.

En outre, les perspectives de nos travaux sont multiples et partagées en perspectives à court terme et d'autres à long terme.

A court terme, nous souhaitons intégrer le détecteur de peau fondé sur l'arbre de dépendance à bi-probabilité peau et non peau Discriminatoire dans un système embarqué à temps réel. Nous choisissons en particulier de concevoir un filtre d'Internet pour la gestion du contenu visuel d'adultes. En particulier nous nous intéressons au filtrage des images d'adultes sur le Web. Le processus de la détection de telles images est principalement basé sur la détection de la peau qui constitue son premier chaînon. Cela est dû au constat qu'une grande zone du corps humain est nue dans ce type des images traitées. A partir du modèle de l'arbre de dépendances à bi-probabilité peau et non peau Discriminatoire qui a présenté les meilleurs performances en termes de qualité de détection, nous proposons un algorithme de filtrage des images d'adultes sur le Web. D'autant plus que le temps réel est une caractéristique pertinente de ce genre d'application, nous envisageons d'exploiter d'avantage la propriété d'homogénéité de la peau pour modifier notre approche de la détection de la peau afin qu'elle puisse répondre à la contrainte du temps.

Vue l'utilisation massive des modèles graphiques dans plusieurs secteurs d'activité, il paraît être intéressant d'exploiter nos modèles dans d'autres domaines de recherches. Il est à noter que les nouveaux modèles de la détection de la peau proposés dans ce mémoire, peuvent être adaptés et utilisés dans la modélisation des distributions de probabilités de deux classes complémentaires. Nous désignons par deux classes complémentaires, deux classes dont l'union constitue l'univers, exemples : peau/non peau, image interne/image externe, image d'adultes/image non adultes, etc. Ainsi à long terme, nous envisageons l'utilisation du modèle d'arbre de dépendance à bi-probabilité pour la classification d'autres classes ou encore la détection des objets dans les images. Il suffit de choisir les descripteurs convenables pour accomplir la tâche demandée. De plus, nos approches conviennent aussi dans le cas de multi-classification. En effet, il suffit d'adopter une démarche dichotomique pour différencier entre toutes les classes. Une telle approche permettra de minimiser le temps de traitement et l'espace mémoire pour la simple raison,

---

qu'un arbre représente deux classes d'où le nombre total d'arbres nécessaire à utiliser et à enregistrer, est égal au nombre de classes utilisées moins un.



---

**LE MÉLANGE DES ARBRES DE CHOW LIU : THÉORÈMES**

Dans le cadre de la recherche d'un arbre recouvrant "idéal" sur un graphe qui minimise la distance de Kullback entre la distribution de probabilité réelle et celle modélisée, l'objectif des deux sous section qui succèdent consiste à prouver deux théorèmes sur le mélange des ArbresROpts de Chow et Liu. Nous notons que les notations définies au chapitre 4 sont conservées.

## A.1 Preuve du théorème 5 chapitre 4

Rappelons le théorème à prouver :

Théorème 5.

*Une distribution de probabilité est mieux modélisée par l'arbre  $T_{mix}$  obtenu par le mélange des arbres recouvrants de l'ensemble  $S^K$ , que par le "plus mauvais arbre" de ce mélange. Autrement écrit :*

$$KL(P(x), P(x|T_{mix})) \leq KL(P(x), P(x|T')) \quad , \quad T' \in S^K \quad (A.1)$$

**Preuve. 1.** .

*Nous comparons les distances de Kullback entre, d'une part la distribution du probabilité réelle et celle modélisée par un arbre de mélange ( $T_{mix}$ ) de  $K$  arbres recouvrants, et d'autre part la distribution du probabilité réelle et celle portée par le "plus mauvais arbre" des  $K$*

arbres considérés ( $T'$ ).

A partir de l'équation (2.8) et par définition de la probabilité du mélange, nous obtenons :

$$KL(P(x), P(x|T_{mix})) = \sum_{x \in V} P(x) \log \left( \frac{P(x)}{P(x|T_{mix})} \right) \quad (\text{A.2})$$

$$= \sum_{x \in V} P(x) \log (P(x)) - \sum_{x \in V} P(x) \log \sum_{T \in \mathcal{S}^K} \lambda(T) P(x|T) \quad (\text{A.3})$$

$$\begin{aligned} &= \sum_{x \in V} P(x) \log (P(x)) - \sum_{x \in V} P(x) \log \left( \prod_{u \in V} P_u(x_u) \right) \\ &\quad - \sum_{x \in V} P(x) \log \sum_{T \in \mathcal{S}^K} \lambda(T) \prod_{(u \sim v) \in T} \frac{P_{uv}(x_u, x_v)}{P_u(x_u) P_v(x_v)} \end{aligned} \quad (\text{A.4})$$

Etant donné que la fonction  $\log$  est une fonction réelle continue et concave et que les coefficients  $\lambda(T)$  sont des poids réels et positifs, ainsi en utilisant l'inverse de l'égalité de Jensen nous aboutissons à :

$$\begin{aligned} &\log \sum_{T \in \mathcal{S}^K} \left( \lambda(T) \prod_{(u \sim v) \in T} \frac{P_{uv}(x_u, x_v)}{P_u(x_u) P_v(x_v)} \right) \\ &\geq \sum_{T \in \mathcal{S}^K} \lambda(T) \log \left( \prod_{(u \sim v) \in T} \frac{P_{uv}(x_u, x_v)}{P_u(x_u) P_v(x_v)} \right) \end{aligned} \quad (\text{A.5})$$

Il suit que

$$\begin{aligned} KL(P(x), P(x|T_{mix})) &\leq \sum_{x \in V} P(x) \log (P(x)) - \sum_{x \in V} P(x) \log \prod_{u \in V} P_u(x_u) \\ &\quad - \sum_{x \in V} P(x) \sum_{T \in \mathcal{S}^K} \lambda(T) \log \left( \prod_{(u \sim v) \in T} \frac{P_{uv}(x_u, x_v)}{P_u(x_u) P_v(x_v)} \right) \end{aligned} \quad (\text{A.6})$$

Or,

$$\begin{aligned} &\sum_{x \in V} P(x) \sum_{T \in \mathcal{S}^K} \lambda(T) \log \left( \prod_{(u \sim v) \in T} \frac{P_{uv}(x_u, x_v)}{P_u(x_u) P_v(x_v)} \right) \\ &= \sum_{T \in \mathcal{S}^K} \lambda(T) \sum_{(u \sim v) \in T} \sum_{x \in V} P(x) \log \left( \frac{P_{uv}(x_u, x_v)}{P_u(x_u) P_v(x_v)} \right) \end{aligned} \quad (\text{A.7})$$

$$= \sum_{T \in \mathcal{S}^K} \lambda(T) \sum_{(u \sim v) \in T} KL(P_{uv}(x_u, x_v), P_u(x_u)P_v(x_v)) \quad (\text{A.8})$$

$$= \sum_{T \in \mathcal{S}^K} \lambda(T) W^T \quad (\text{A.9})$$

Considérons  $T'$  le plus mauvais arbre de l'ensemble des arbres  $\mathcal{S}^K$  et qui vérifie l'équation (4.7), nous avons :

$$W^{T'} \leq W^T \quad , \quad \forall T \in \mathcal{S}^K \quad (\text{A.10})$$

De plus,

$$\sum_{T \in \mathcal{S}^K} \lambda(T) = 1 \quad , \quad \lambda(T) \geq 0 \quad (\text{A.11})$$

Donc

$$\sum_{T \in \mathcal{S}^K} \lambda(T) W^T \geq W^{T'} = \sum_{(u \sim v) \in T'} KL(P_{uv}(x_u, x_v), P_u(x_u)P_v(x_v)) \quad (\text{A.12})$$

D'où

$$\begin{aligned} KL(P(x), P(x|T_{mix})) &\leq \sum_{x \in V} P(x) \log(P(x)) - \sum_{x \in V} P(x) \log\left(\prod_{u \in V} P_u(x_u)\right) \\ &\quad - \sum_{(u \sim v) \in T'} KL(P_{uv}(x_u, x_v), P_u(x_u)P_v(x_v)) \end{aligned} \quad (\text{A.13})$$

Il suit que,

$$KL(P(x), P(x|T_{mix})) \leq KL(P(x), P(x|T')) \quad (\text{A.14})$$

Preuve conclue.

## A.2 Preuve du théorème 6 chapitre 4

Il s'agit de prouver le théorème suivant :

### Théorème 6.

La meilleure probabilité modélisée par un mélange d'arbres recouvrants du graphe  $G$ , est obtenue par l'ensemble de  $K$  arbres recouvrants optimaux de celui-ci.

**Preuve. 2. .**

Ce théorème est exprimé formellement par l'inégalité (4.12). Or nous avons :

$$\begin{aligned}
KL(P(x), P(x|T_{mix}^{\Theta^K})) &= \sum_{x \in V} P(x) \log(P(x)) - \sum_{x \in V} P(x) \log \left( \prod_{u \in V} P_u(x_u) \right) \\
&\quad - \sum_{x \in V} P(x) \log \sum_{T \in \Theta^K} \lambda(T) \prod_{(u \sim v) \in T} \frac{P_{uv}(x_u, x_v)}{P_u(x_u)P_v(x_v)} \quad (A.15)
\end{aligned}$$

En utilisant l'inverse de l'inégalité de Jensen, nous obtenons :

$$\begin{aligned}
KL(P(x), P(x|T_{mix}^{\Theta^K})) &\leq \sum_{x \in V} P(x) \log(P(x)) - \sum_{x \in V} P(x) \log \prod_{u \in V} P_u(x_u) \\
&\quad - \sum_{x \in V} P(x) \sum_{T \in \Theta^K} \lambda(T) \log \left( \prod_{(u \sim v) \in T} \frac{P_{uv}(x_u, x_v)}{P_u(x_u)P_v(x_v)} \right) \quad (A.16)
\end{aligned}$$

Aussi :

$$\sum_{x \in V} P(x) \sum_{T \in \Theta^K} \lambda(T) \log \left( \prod_{(u \sim v) \in T} \frac{P_{uv}(x_u, x_v)}{P_u(x_u)P_v(x_v)} \right) = \sum_{T \in \Theta^K} \lambda(T) W^T \quad (A.17)$$

De l'ensemble des ArbresROpts  $\Theta^K$ , nous considérons un optimal arbre recouvrant  $T^{op}$ .

Par définition nous avons ces quatre formules :

$$W^{T^{op}} = W^T \quad , \quad \forall T \in \Theta^K \quad (A.18)$$

$$\sum_{T \in \Theta^K} \lambda(T) = 1 \quad , \quad \lambda(T) \geq 0 \quad (A.19)$$

$$W^{T^{op}} \geq W^{T^s} \quad , \quad \forall T^{op} \in \Theta^K \quad \forall T^s \in S^K \quad (A.20)$$

$$\sum_{T \in S^K} \lambda'(T) = 1 \quad , \quad \lambda'(T) \geq 0 \quad (A.21)$$

Il suit donc :

$$\sum_{T^s \in S^K} \lambda'(T^s) W^{T^{op}} \geq \sum_{T^s \in S^K} \lambda'(T^s) W^{T^s} \quad (A.22)$$

*d'où*

$$W^{T^{op}} \geq \sum_{T^s \in S^K} \lambda'(T^s) W^{T^s} \quad (\text{A.23})$$

*Finalement*

$$KL(P(x), P(x|T_{mix}^{\Theta^K})) \leq KL(P(x), P(x|T_{mix}^{S^K})) \quad (\text{A.24})$$

*Preuve conclue.*





---

## BIBLIOGRAPHIE

- [1] H. Ouhaddi, P. Horain, K. Milkolajczyk. Modélisation et suivi de la main. Actes 4èmes Journées d'études et d'échanges Compression et REprésentation des Signaux Audiovisuels (CORESA'98), Lannion, France, 9-10 June 1998, pp. 109-114, 1998.
- [2] J.S. Pierrard and T. Vetter. Skin Detail Analysis for Face Recognition. IEEE Conference on Computer Vision and Pattern Recognition. June 17-22 2007. Pages : 1-8.
- [3] 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.
- [4] S. Mitra and T. Acharya. Gesture Recognition : A Survey. IEEE Transactions on Systems, Man and Cybernetics, Part C : Applications and Reviews. Volume 37, Issue 3, May 2007, Pages : 311-324.
- [5] A. Ramamoorthy, N. Vaswani, S. Chaudhury, and S. Banerjee. Recognition of dynamic hand gestures. Pattern Recognition. Volume 36, Issue 9, Septembre 2003, Papes : 2069-2081.
- [6] A. Jacquin and A. Eleftheriadis, "Automatic location tracking of faces and facial features in video signal", Int. Work. On Automatic Face and Gesture Recognition, Pages : 142-147, Zurich, 1995.
- [7] K. Aas, "Detection and recognition faces in video sequences", Norsk regnesentral, 1998.

- [8] J.S. Lee, Y.M. Kuo, P.C. Chung, and E.L. Chen. Naked image detection based on adaptive and extensible skin color model. *Pattern Recognition*. Volume 40, Issue 8, August 2007, Pages : 2261-2270.
- [9] H. Zheng, M. Daoudi, and B. Jedynek. Blocking Adult Images Based on Statistical Skin Detection. *Electronic Letters on Computer Vision and Image Analysis*. Volume 4, Number 2, 2004, Pages : 1-14.
- [10] L. Sigal, S. Sclaroff, and V. Athitsos. Skin color-based video segmentation under time-varying illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Pages : 862-877, July 2004.
- [11] Gu. Lifang and D. Bone. Skin colour region detection in MPEG video sequences. *International Conference on Image Analysis and Processing*. September 27-29 1999, Venice, Italy. Pages : 898-903.
- [12] 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.
- [13] D. A. Forsyth and M. M. Fleck. Identifying nude pictures. *IEEE Workshop on the Applications of Computer Vision*, pp. 103-108, 1996.
- [14] 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.
- [15] H. Zheng. Modèles de maximum d'entropie pour la détection de la peau : application au filtrage de l'internet. Thèse d'état., Université de Sciences et Technologies de Lille, Lille, France, Novembre 2004.
- [16] 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*, volume 1, pages 1056-1059, Barcelona, Spain, September 2000.
- [17] M.J. Jones and J. M. Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1) :81-96, January 2002.
- [18] 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, pages 491-500, 2001.
- [19] L.M. Bergasa, M. Mazo, A. Gardel, M.A. Sotelo, L. Boquete, Unsupervised and adaptive Gaussian skin-color model, *Image Vision Comput.* 18 (12) (2000) 987-1003.

- [20] N. Sebe, T. Cohen, T.S. Huang, T. Gevers, Skin detection, a Bayesian network approach, ICPR04, 2004.
- [21] D. Brown, I. Craw, J. Lewthwaite, A SOM based approach to skin detection with application in real time systems, BMVC01, 2001.
- [22] Q.H. Thu, M. Meguro, M. Kaneko, Skin-color extraction in images with complex background and varying illumination, Sixth IEEE Workshop on Applications of Computer Vision, 2002.
- [23] L. Sigal, S. Sclaroff, and V. Athitsos. Skin color-based video segmentation under time-varying illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(7) :862-877, July 2004.
- [24] Q. Zhu, K.-T. Cheng, C.-T. Wu, Y.-L. Wu, Adaptive learning of an accurate skin-color model, AFGR04, 2004.
- [25] B. S. Manjunath, Philippe Salembier et Thomas Sikora, éditeurs. *Introduction to MPEG-7 : Multimedia Content Description Language*. Wiley-Interscience, 2002.
- [26] 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.
- [27] 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.
- [28] J. Ahlberg, *Extraction and Coding of Face Model Paramaters*, Licentiate Thesis No. 747, Departement of Electrical Engineering, Linköping University, Sweden, March 1999.
- [29] 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) :696-706, May 2002.
- [30] D. Chai and A. Bouzerdoum. A bayesian approach to skin color classification in ycbcr color space. In *Proc. IEEE Region Ten Conference(TENCON'2000)*, volume 2, pages 421-424, 2000.
- [31] Jean-Christophe Terrillon, M. N. Shirazi, H. Fukamachi, and S. Akamatsu. Comparative performance of different skin chrominance models and chrominance spaces

- for the automatic detection of human faces in color images. In Fourth International Conference On Automatic Face and gesture Recognition, pages 54-61, 2000.
- [32] Y. Dai, Y. Nakano, Face-texture model based on SGLD and its application in face detection in a color scene, *Pattern Recognition* 29 (6) (1996) 1007-1017.
- [33] 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*, volume 1, pages 1056 - 1059, Barcelona, Spain, September 2000.
- [34] 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.
- [35] J. Cai, A. Goshtasby, Detecting human faces in color images, *Image Vision Comput.* 18 (1999) 63-75.
- [36] S. Kawato, J. Ohya, Automatic skin-color distribution extraction for face detection and tracking, *Fifth International Conference on Signal Processing*, vol. 2, 2000, pp. 1415-1418.
- [37] M.H. Yang, N. Ahuja, Gaussian Mixture model for human skin color and its application in image and video databases, *Proceedings of SPIE : Conference on Storage and Retrieval for Image and Video Databases*, vol. 3656, 1999, pp. 458-466.
- [38] 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.
- [39] 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*, pages 58-63, Corfu, Greece, September 1999.
- [40] A. Albiol, L. Torres, and E.J. Delp. Optimum color spaces for skin detection. In *Proc. of the International Conference on Image Processing*, volume 1, pages 122- 124, Tesseloniki, Greece, 2001.
- [41] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. A Survey of Skin-Color Modeling and Detection Methods. *Pattern Recognition*. Volume 40, Issue 3, March 2007, Pages : 1106-1122.

- [42] V. Vezhnevets, V. Sazonov, A. Andreeva, A survey on pixel-based skin color detection techniques, GRAPHICON03, 2003, pp. 85 - 92.
- [43] Zhanwu Xu and Miaoliang Zhu. Color-Based Skin Detection : Survey and Evaluation. Twelfth International Multi-Media Modelling Conference. January 4-6 2006. Pages : 143-152.
- [44] 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, pages 58 - 63, Corfu, Greece, September 1999.
- [45] Jean-Christophe Terrillon, M. N. Shirazi, H. Fukamachi, and S. Akamatsu. Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In Fourth International Conference On Automatic Face and gesture Recognition, pages 54 - 61, 2000.
- [46] 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.
- [47] 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.
- [48] 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.
- [49] F. Bérard, « Vision par ordinateur pour l'interaction homme-machine fortement couplée », thèse de doctorat, Université Joseph Fourier, Grenoble, France, Janvier 2000.
- [50] 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.
- [51] F. K. H. Quek, T. Mysliwiec, and M. Zhao. Fingermouse : A freehand pointing interface. Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition, Zurich, Switzerland, pp. 372-377, June 1995.

- [52] M.J. Jones and J. M. Rehg. Statistical color models with application to skin detection, *International Journal of Computer Vision*. Volume 46, Issue 1, January 2002, Pages : 81-96.
- [53] 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.
- [54] G. Gomez, M. Sanchez, L.E. Sucar, On selecting an appropriate colour space for skin detection, *Springer-Verlag : Lecture Notes in Artificial Intelligence*, vol. 2313, 2002, pp. 70-79.
- [55] S. Marcel, and S. Bengio, Improving face verification using skin color information, *ICPR02*, 2002.
- [56] M.M. Fleck, D.A. Forsyth, C. Bregler, Finding naked people, *Proceedings of European Conference on Computer Vision*, vol. 2, 1996, pp. 592-602.
- [57] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, *Mach. Learn.* 29 (2) (1997) 131-163.
- [58] H. Sahbi, N. Boujemaa, Coarse to fine face detection based on skin color adaptation, *Workshop on Biometric Authentication*, 2002, *Lecture Notes in Computer Science*, vol. 2359, 2002, pp. 112-120.
- [59] M.J. Seow, D. Valaparla, V.K. Asari, Neural network-based skin color model for face detection, *Proceedings of the 32nd Workshop on Applied Imagery Pattern Recognition*, 2003.
- [60] J. Karlekar, U.B. Desai, Finding faces in color images using wavelet transform, *International Conference on Image Analysis and Processing*, 1999.
- [61] S.L. Phung, D. Chai, A. Bouzerdoum, A universal and robust human skin color model using neural networks, *IJCNN01*, 2001.
- [62] C. Chen, S.P. Chiang, Detection of human faces in colour images, *IEE Proc. Vision Image Signal Process.* 144 (6) (1997) 384-388.
- [63] T. Kohonen, "Self-Organized Formation of Topologically Correct Feature Maps", *Biological Cybernetics*, vol. 46, pp. 59-69, 1982.

- [64] T. Kohonen, *Self-Organizing Maps*, vol. 30. Berlin : Springer-Verlag, 1995.
- [65] M. Talibi Alaoui, R. Touahni et A. Sbihi, Classification des Images Couleurs par association des Transformations Morphologiques aux Cartes de Kohonen, CARI 2004, pp.83-90.
- [66] 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.
- [67] E. Saber and A. Tekalp. Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions. *Pattern Recognition Letters*, 19(8), pp. 669-680, 1998.
- [68] M. Yang and N. Ahuja. Detecting human faces in color images. *International Conference on Image Processing (ICIP)*, Chicago, pp.127- 130, 1998.
- [69] 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.
- [70] J. C. Terrillon, M. N. Shirazi, H. Fukamachi, and S. Akamatsu. Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In *Fourth International Conference on Automatic Face and gesture Recognition*, pp. 54-61, 2000.
- [71] M. Hunke and A. Waibel, "Face locating and tracking for humain computer interaction", *IEEE Computer*, pp. 1277-1281, November 1994.
- [72] 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.
- [73] 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*.
- [74] Abdullah-Al-Wadud, M. Oksam Chae. "Region-of-Interest Selection for Skin Detection Based Applications". *International Conference on Convergence Information Technology*. Pages : 1999-2004. 21-23 November 2007.
- [75] J. Kovac, P. Peer, F. Solina, "2D Versus 3D Colour Space Face Detection", *4th EURASIP Conference on Video/Image Processing and Multimedia Communications, Croatia, 2003*, pp. 449-454.

- J. Kovač, P. Peer, and F. Solina. Human skin colour clustering for face detection. In B. Zajc, editor, International Conference on Computer as a Tool, EUROCON, Ljubljana, Slovenia, September 2003.
- [76] F. Gasparini, R. Schettini, "Skin Segmentation using Multiple Thresholding", SPIE proceedings, Vol. 6061-18, January 2006.
- [77] 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.
- [78] I. Craw, D. Tock et A. Bennett, "Finding face features", Proc. 2nd European Conference on computer Vision, pp. 92-96, 1992.
- [79] 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.
- [80] 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.
- [81] 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.
- [82] D. Chetverikov and A. Lerch, "Multiresolution Face Detection", Theoretical Foundations of Computer Vision, vol. 69, Pages : 131 - 140, 1993.
- [83] K. C Yow and R. Cipolla, "Finding initial estimates of humain face location", Department of Engineering, university of Cambridge, 1995.
- [84] A. Lemieux and M. Parizeau, Flexible multi-classifier architecture for face recognition systems. Vision Interface, 2003.
- [85] T. Horprasert, D. Harwood and L. Davis, A robust background subtraction and shadow detection. In Proceedings of the ACCV, 2000.
- [86] 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.
- [87] 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.

- [88] 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.
- [89] 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.
- [90] V. Govindaraju, D. B. Sher, R.K. Srihari et S.N. Srihari, "locating human faces in newspaper photographs", Proc. Conf. On Comp. Vision and Pattgern Recognition, pp. 549-554, 1989.
- [91] B. Jedynak, H. Zheng, and M. Daoudi. Skin Detection Using Pairwise Models. Image and Vision Computing. Volume 23, Issue 13, 29 November 2005, Pages : 1122-1130.
- [92] C. Chow and C. Liu. Approximating Discrete Probability Distributions with Dependence Trees. Fifteenth IEEE Transactions on Information Theory. Volume 14, Issue 3, May 1968, Pages : 462-467.
- [93] ChihsinWu and Peter C. Doerschuk. Tree approximations to markov random fields. IEEE Trans. on PAMI, 17(4) :391 - 402, April 1995.
- [94] T. M. Cover and J. A. Thomas. Elements of Information Theory. Wiley-Interscience, August 1991.
- [95] J. Pearl. Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference. Morgan Kaufmann, 1988.
- [96] N. Friedman, O. Mosenzon, N. Slonim, and N. Tishby. Multivariate Information Bottleneck. Seventeenth Conference on Uncertainty in Artificial Intelligence, Seattle, Washington, USA, August 2-5 2001. Pages : 144-151.
- [97] M. Meila and M.I. Jordan. Learning with Mixtures of Trees, Journal of Machine Learning Research, Volume 1, Pages : 1-48, 2000.
- [98] A. Cayley. A theorem on trees. Mathematical Papers, 13 :26 - 28, 1897
- [99] N. Katoh, T. Ibaraki, and H. Mines. An Algorithm for Finding K Minimum Spanning Trees, In Society for Industrial and Applied Mathematics, Volume 10, Number 2, May 1981.
- [100] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm (with discussion). Journal of the Royal Statistical Society, B 39 :1 - 38. 1977.

- 
- [101] M. C. Fairhurst and A. F. R. Rahman. Introducing New Multiple Expert Decision Combination Topology, : A Case Study Using Recognition of Handwritten Characters, Proc. of ICDAR'97, Pages :886-891, Ulm, 1997.
- [102] B. Moobed. Combinaison de classifieurs, une nouvelle approche, Thèse de Doctorat de l'Université de Paris Sud - Orsay, 1996.
- [103] B. Jedynek, H. Zheng, and M. Daoudi. Skin Detection Using Pairwise Models. Image and Vision Computing. Volume 23, Issue 13, 29 November 2005, Pages : 1122-1130.