

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mouloud MAMMERY, Tizi-Ouzou



Faculté de Génie Electrique et d'Informatique
Département d'Automatique

Mémoire de Fin d'Etudes

En vue de l'obtention du diplôme de

Master académique en automatique

Option commande des systèmes

Thème

***Réseaux bayésiens en traitement
d'images***

Application à la détection de la peau

Dirigé par :

M. DIAF.M

Présenté par :

M^{elle} MAMERY Liza

M^{elle} LARABI Chafia

Soutenu le : 07 / 10 /2012

Promotion 2012

REMERCIEMENTS

Nous tenons à remercier en premier lieu, le Bon Dieu qui nous a montré le bon chemin et nous a donné le courage et la volonté à mener à terme ce Travail, et l'achever avec beaucoup de patience.

Nous exprimons nos respects et remerciements à Monsieur DIAF Moussa, professeur à L'UMMTO de nous avoir fait confiance, en proposant et dirigeant ce travail.

Nos sincères et vifs remerciements s'adressent aussi aux membres du jury qui ont accepté d'examiner notre travail.

Un remerciement particulier s'adresse à tous les étudiants de la faculté de génie électrique et d'informatique, surtout aux étudiants du MASTER 2, AUTOMATIQUE ACADEMIQUE et qui nous ont aidé en se montrant volontiers à prendre des photos afin de construire notre base d'images.

Enfin, il nous serait difficile d'omettre de remercier tous ceux qui ont contribué de près ou de loin à réaliser ce travail. Qu'ils trouvent dans ses quelques lignes l'expression de nos sincères remerciements.

L. MAMERI

C. LARABI

A

nos parents ;

nos frères et sœurs ;

nos amis et tous ceux qui nous sont chers.

« S'il n'y avait pas d'hiver, le printemps ne serait pas si agréable : Si nous ne goutons pas à l'adversité, la réussite ne serait pas toute appréciée. »

Anne Bradstreet

Résumé

Les réseaux bayésiens sont un outil qui s'adapte non seulement au raisonnement avec la connaissance et la croyance incertaine à partir de données incomplètes, mais aussi à la structure de la représentation de la connaissance. Leur succès est fortement lié à leur capacité de représenter et de manipuler des relations de (in) dépendance qui sont importantes pour une gestion efficace des informations incertaines. Dans notre manuscrit, nous appliquons les réseaux bayésiens à la détection de la peau dans des images couleurs. A cet effet, nous avons opté pour le réseau bayésien naïf (NBN) et cela en classifiant les pixels peau et les pixels non peau. Nous montrons que ce classifieur bayésien donne des résultats probants malgré leur structure simplifiée.

Mots clés : Réseaux bayésiens, Réseaux bayésiens naïfs, Détection de la peau, Apprentissage, Inférence.

Abstract

The bayesian networks are an adaptable means not only to reasoning with the knowledge and the uncertain from incomplete data but to the structure of the knowledge representation as well their success is strongly linked to their ability to represent and manipulate (in) dependence relations which are important for an efficient dealing with uncertain data. In our manuscript, we apply the bayesian networks to the skin detection in colour images. For this, we applied the Naïve bayesian networks (NBN) by classifying skin pixels and non skin pixels. We show that this bayesian classifier gives convincing results despite their simplified structure.

Key words : Bayesian networks, Naïve Bayesian networks, Skin detection, Learning, Inference.

Sommaire

Introduction générale.....	1
-----------------------------------	----------

Chapitre 1 Notions sur le traitement d'images

1. Introduction.....	4
2. Prétraitement d'une image.....	5
3. Quelques opérations de traitement d'images.....	7
3.1 L'amélioration d'images.....	7
3.2 La restauration d'images.....	8
3.3 La segmentation d'images	8
3.3.1 Segmentation par approche régions	9
3.3.2 Segmentation par approche contour	10
3.3.3 Segmentation par classification	10
4. Règle de décision de Bayes et réseaux bayésiens	10
4.1 Décision bayésienne.....	11
4.2 Réseaux bayésiens	11
5. La couleur en traitement d'images.....	12
6. Conclusion.....	17

Chapitre 2 Généralités sur les réseaux bayésiens

1. Introduction.....	18
2. Représentation graphique de la causalité	19
2.1 Circulation de l'information dans un graphe causal	19
2.2 Différents types de connexion	20
2.3 Notion de d-séparation.....	21
3. Représentation probabiliste associée	21
3.1 Indépendance conditionnelle.....	21
3.2 Théorème de Bayes	22
4. Réseaux bayésiens	23
5. L'inférence dans les réseaux bayésiens	24
5.1 Les méthodes exactes	24
5.1.1 Propagation de messages locaux dans un arbre	24
5.1.2 Inférence dans une structure générique: Algorithme de Jtree	27

5.1.3	Conditionnement global.....	28
5.1.4	Algorithme d'Elimination de Bucket.....	29
5.2	Méthodes d'inférence approchées	29
6	Apprentissage dans les réseaux bayésiens	30
6.1	Apprentissage des paramètres.....	30
6.1.1	Apprentissage à partir de données complètes.....	30
6.1.2	Apprentissage à partir de données incomplètes.....	31
6.2	Apprentissage de la structure	32
6.2.1	Méthodes basées sur les contraintes et recherche d'indépendances	32
6.2.2	<i>Méthodes basées sur l'optimisation d'un score.....</i>	33
6.2.3	<i>Méthodes hybrides.....</i>	33
6.2.4	Quelques algorithmes d'apprentissage de la structure...33	
7.	Extensions de réseaux bayésiens34	
7.1	Réseaux bayésiens multi-agents	34
7.2	Réseaux bayésiens orientés objets	34
7.3	Diagrammes d'influence	35
7.4	Réseaux bayésiens dynamiques.....	35
7.5	Les réseaux bayésiens adaptés à la classification.....	36
7.5.1	Les réseaux bayésiens naïfs.....	36
7.5.2	Réseaux bayésiens naïfs augmentés (TANB).....	37
7.5.3	Les multi-nets bayésiens	37
7.6	Réseaux bayésiens causaux.....	38
8.	Réseaux bayésiens à densité de probabilité continue	39
9.	Conclusion	40

Chapitre 3 Détection de la peau

1.	Introduction	41
2.	Les méthodes de détection de la peau.....	42
2.1	Approche basée sur la couleur.....	42
2.1.1	Les méthodes basées sur des seuils fixés de définition de la peau.....	43

2.1.2 Méthodes paramétriques	44
2.1.3 Méthodes non paramétriques	45
3. Classification par la règle de Bayes.....	47
3.1 Classifieur bayésien couplé à un histogramme.....	47
3.2 Classification par réseau bayésien.....	48
4. Tests et résultats.....	49
4.1 Méthode adoptée	51
4.2 Discussion des résultats.....	57
5. Conclusion.....	58
Conclusion générale et perspectives	59

Listes des figures

Premier chapitre : Notions sur le traitement d'image

1. Filtres passe-haut et passe-bas
2. Filtre médian
3. Courbe tonale
4. Les courbes d'appariement $R(\lambda)$, $V(\lambda)$ et $B(\lambda)$
5. Cube des couleurs dans le modèle triangulaire
6. Modèle du cône hexagonal

Deuxième chapitre : Généralités sur les réseaux Bayésiens

1. représentation graphique de la causalité
2. modélisation de l'exemple de la pelouse mouillée
3. Connexion convergente
4. Connexion en série
5. Connexion divergente
6. Processus de fonctionnement d'un réseau bayésien
7. Mécanisme de propagation de connaissances au niveau d'un nœud dans un réseau bayésien à structure d'arbre
8. Etape de l'algorithme Jtree. (a) graphe original, (b) graphe moralisé, (c) graphe triangulé, (d) arbre de jonction
9. Exemple de réseau bayésien multi agents Exemple de réseau bayésien multi agents (RBMA)
10. Exemple d'un diagramme d'influence
11. HMM sous forme de réseau bayésien
12. Réseau bayésien naïf
13. Réseau bayésien naïf augmenté
14. différents types de modélisation par variables latentes

Troisième chapitre : Détection de la peau par Réseau bayésien naïf

1. Diversité ethniques
2. Différents éclairages
3. Décors complexes
4. Disposition des neurones de Kohonen
5. principe de détection de peau
6. exemple des images d'apprentissage
7. Réseau bayésien naïf obtenu avec les paramètres d'intensités dans le rouge, le vert et le bleu
8. Détection de la peau à partir d'une image RGB
9. Détection de la peau à partir d'une image YCbCr
10. Détection de la peau à partir d'une image HSV

Liste des tableaux

1. les variables de l'exemple de la pelouse mouillée
2. Résultats obtenus en appliquant le RBN et les seuils pour la détection de la peau à partir des images RGB
3. résultats obtenus en appliquant le RBN et les seuils pour la détection de la peau à partir des images YCbCr
4. résultats obtenus en appliquant le RBN et les seuils pour la détection de la peau à partir des images HSV.

Introduction générale

L'information n'est pas la connaissance. A mesure que se développent les technologies permettant de stocker, d'échanger de l'information et d'y accéder, la question de l'analyse et de synthèse de ces informations devient essentielle. Le développement de technologies facilitant le passage de l'information à la connaissance est déterminant pour que la société annoncée de l'information ne soit pas en réalité la société du bruit.

Les modèles graphiques probabilistes, et plus précisément les réseaux bayésiens, sont le résultat d'une convergence entre deux types d'approches : les méthodes statistiques conçues pour permettre le passage de l'observation à la loi de probabilité, les technologies de l'intelligence artificielle dont l'objectif est de permettre aux ordinateurs de traiter de la connaissance plutôt que de l'information. Ils constituent aujourd'hui l'un des formalismes les plus complets et les plus cohérents pour l'acquisition, la représentation et l'utilisation de la connaissance par des ordinateurs. Encore du domaine de la recherche au début des années 1990, cette technologie connaît de plus en plus d'applications.

Toute application mettant en œuvre des connaissances peut relever de l'utilisation des réseaux bayésiens du fait qu'ils constituent une technique d'acquisition, de représentation et de manipulation de connaissances, qu'il s'agisse de formaliser la connaissance d'expert, d'extraire la connaissance contenue dans des bases de données, ou d'utiliser le plus rationnellement possible l'un ou l'autre type de connaissance. Les réseaux bayésiens sont des modèles de représentation d'informations incertaines et de raisonnement à partir d'informations incomplètes sur des phénomènes complexes.

Il faut noter que l'appellation « réseaux bayésiens » est prête à confusion. En effet, ceux-ci ne sont pas forcément des modèles bayésiens au sens statique du terme. Ce sont des modèles graphiques probabilistes utilisant le théorème de Bayes pour raisonner. Conçus pour pouvoir prendre en charge les problèmes comportant la notion d'incertitude, les réseaux bayésiens apportent à la fois une interface intuitive sous la forme d'un graphe orienté et un ensemble de méthodes permettant d'exploiter au mieux la connaissance extraite qu'ils modélisent. Par conséquent, les réseaux bayésiens se sont peu à peu imposés parmi les différents modèles probabilistes existants.

Les réseaux bayésiens étant des modèles graphiques probabilistes bénéficient non seulement des avantages des modèles probabilistes, mais aussi, ils présentent des avantages supplémentaires liés à leur représentation graphique. Cependant, beaucoup d'expériences montrent qu'il est souvent plus facile de formaliser les connaissances sous forme d'un graphe que sous forme d'un système basé sur les règles. En outre, la modélisation d'un problème par un réseau bayésien, puis l'utilisation d'algorithmes d'inférence ont fait des réseaux bayésiens des outils idéaux pour le raisonnement à partir d'informations incomplètes.

Les applications des réseaux bayésiens peuvent être classées par type ou par domaine. Dans le premier cas, les applications reposent en général sur deux types de modèles : les modèles symboliques pour le diagnostic et la planification et les modèles numériques pour la classification, la prévision et le contrôle. Dans le deuxième cas, les domaines d'application des réseaux bayésiens sont très vastes : on retrouve le domaine de la santé, l'industrie, la défense, l'informatique et la sécurité.

L'application que nous avons abordée dans notre travail est utilisée dans le domaine de la sécurité. Il s'agit de la détection de la peau humaine dans des images couleur. Ainsi, nous avons opté pour l'utilisation d'un réseau bayésien destiné à la classification appelé « *Réseau bayésien naïf* » dans le but de distinguer les pixels de peau de ceux de non-peau. La classification comme la segmentation, l'extraction, la recherche, l'indexation, est une manière d'analyser le contenu visuel d'une image qui reste le moyen le plus efficace pour le traitement d'une image. Les méthodes de classification se distinguent par leur caractère supervisé et non supervisé. Il y a aussi des auteurs qui parlent de classification semi-supervisée qui n'est autre qu'une combinaison entre les deux premiers types de classification cités.

La détection de la peau est un sujet de recherche d'un enjeu important dans la mesure où il est indispensable d'envisager des analyses et des traitements de niveau supérieur. Il est d'une utilité potentielle pour de nombreuses applications telles que les applications liées à la sécurité (surveillance, contrôle d'accès...). La littérature étale plusieurs techniques de détection de la peau : certaines s'appuient sur les traits caractéristiques, d'autres sur le mouvement et d'autres sur la couleur. Nous nous intéressons à la dernière méthode du fait que l'information couleur reste le moyen le plus couramment utilisé à cette fin.

Chapitre 1

Notions sur le traitement d'images

1. Introduction

On désigne par le traitement d'images numériques l'ensemble des techniques permettant de modifier une image numérique dans le but de l'améliorer ou d'en extraire des informations. Le traitement d'images numériques est une discipline nouvelle qui s'est développée rapidement grâce à l'émergence de nouvelles technologies de l'information. Il s'appuie notamment sur les mathématiques liées à l'information, le traitement du signal, les systèmes électroniques, et sur l'avancée des capacités de calcul des microprocesseurs, notamment ceux qui sont développés exclusivement pour le traitement de signal et qui offrent de grandes capacités et vitesses de calculs. Depuis cinquante ans, un grand nombre d'applications de traitement d'images a été produit dans des domaines très divers. On peut notamment citer l'imagerie médicale, la biométrie, l'asservissement visuel, la robotique, la reconnaissance et le tri d'objets, la télédétection, etc. Dans le processus de traitement d'images, interviennent les moyens matériels et plusieurs opérations telles le prétraitement, le traitement numérique

proprement dit et l'interprétation. Les moyens matériels sont composés principalement de l'unité informatique, du système d'acquisition des images tels les caméras, des moyens d'impression et de stockage. Parmi les opérations les plus usuelles en traitement d'images, on peut citer le prétraitement, la restauration, l'amélioration, la compression, la reconstruction et la segmentation.

2. Prétraitement d'une image

Le prétraitement est une opération fondamentale en traitement d'images. Il permet de compenser certains défauts du capteur, d'améliorer la perception de certains détails, de réduire le bruit et aussi, pour faciliter l'analyse d'une image en renforçant la ressemblance entre pixels appartenant à une même région ou en accentuant la dissemblance entre pixels appartenant à des régions différentes. Le prétraitement est nécessaire notamment dans les cas où l'image est bruitée, le contraste n'est pas suffisant et l'éclairage de l'image n'est pas uniforme. Les prétraitements les plus utilisés sont le lissage local (ou filtrage), l'amélioration d'images et la restauration d'images.

Le lissage local ou filtrage consiste à supprimer le bruit, ou les petites variations présentes dans une image en étudiant pour chaque pixel, les valeurs d'intensité sur son voisinage. En effet, le bruit d'images est un phénomène de brusque variation d'un pixel par rapport à ses voisins. Le filtrage de l'image consiste à modifier la distribution fréquentielle des composantes d'un signal selon des spécifications données.

Deux grandes classes de filtres peuvent être envisagées à savoir, les filtres linéaires et les filtres non linéaires. Les filtres linéaires effectuent des opérations linéaires sur des comptes numériques qui, en général, sont ceux des pixels du voisinage du pixel traité. Ce sont donc des convolutions. On peut citer notamment les filtres passe-bas et les filtres passe-haut.

Filtres passe-bas

Les filtres passe-bas ont un rôle de lissage du signal. Ils adoucissent les variations brutales d'intensité. Comme ils éliminent les fréquences élevées (le bruit), ils modifient assez peu l'image et gardent généralement intacte l'intensité moyenne de l'image. Pour réaliser ce genre de filtrage, on effectue le produit de convolution de l'image par une fonction de voisinage définie d'une façon générale par :

$$H_b = \left[\frac{1}{b+2} \right]^2 \begin{bmatrix} 1 & b & 1 \\ b & b^2 & b \\ 1 & b & 1 \end{bmatrix} \quad (1)$$

où $\left[\frac{1}{b+2} \right]^2$ est un facteur de normalisation

Si $b=1$ est un filtre moyenneur, il prend directement la moyenne algébrique de tous les pixels avec des poids identiques. Si $b=2$, il s'agit d'un filtre passe-bas le plus utilisé. On affecte le poids le plus grand par rapport à son voisinage. De façon générale, les filtres passe-bas sont caractérisés par des coefficients de fenêtre positifs avec un coefficient de valeur maximale pour le point central.

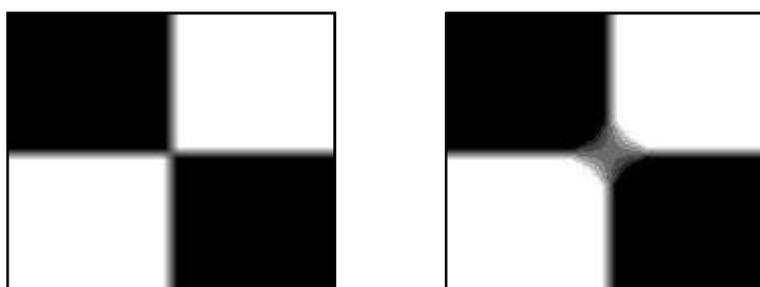
Filtre passe-haut

Les filtres passe-haut ont pour but de rehausser les frontières : ils mettent en valeur les points isolés ou les limites des objets, alors que les autres points sont mis à 0 (en noir). Ceci permet d'accentuer les contrastes, les reliefs et tend à supprimer le flou de l'image. Comme dans le cas de filtre passe-bas, on peut attribuer à chaque point du voisinage d'une fenêtre passe-haut, des poids statistiques différents.



Fig.1. Filtrage passe-haut et passe-bas. (a) Image Léna d'origine. (b) Image après filtrage passe-haut. (c) Image après filtrage passe-bas.

Contrairement aux filtres linéaires qui présentent certains défauts comme le filtre moyenneur qui tend à rendre floues les frontières de l'image traitée, les filtres non linéaires peuvent éviter ces problèmes et peuvent homogénéiser l'image tout en préservant ses frontières. Parmi ces filtres, on peut citer le filtre médian qui ne moyenne plus sur le voisinage, mais on prend la valeur médiane sur ce voisinage. Ce filtre médian est robuste puisque la présence d'un faible nombre de pixels associés à du bruit affecte peu ou pas l'image "résultat". Un des inconvénients de ce filtre est de nécessiter des temps de calcul supérieurs à ceux des filtres linéaires, en raison des nombreux tests informatiques obligatoires. La figure (2) représente un exemple de filtre médian.



(a) (b)

Fig.2 Filtre médian. (a) image avec un flou ; (b) résultat du filtre médian 31×31

3. Quelques opérations de traitement d'images

Le traitement numérique d'images regroupe plusieurs disciplines que l'on classe en deux catégories, à savoir, les traitements de bas niveau et les traitements de haut niveau. Les traitements de bas niveau nécessitent très peu d'informations sur le contenu des images. Ils opèrent en général sur les grandeurs calculées à partir des valeurs attachées à chaque point de l'image sans faire la liaison avec la réalité. Ils opèrent plutôt sur des données de nature numérique. Les traitements de haut niveau, quant à eux, fonctionnent en aval de ceux de bas niveau. Ils peuvent nécessiter des informations sur le contenu des images. Ces traitements s'appliquent sur des entités de natures symboliques associées à une représentation de la réalité extraite de l'image. Dans ce qui suit, nous mentionnons les techniques les plus utilisées en traitement d'images.

3.1 L'amélioration d'images

L'amélioration a pour but d'adapter au mieux les données à une exploitation visuelle par modification de l'information au niveau signal en utilisant uniquement les valeurs de l'image d'entrée. Il peut s'agir de rehausser les contrastes, d'accentuer certaines intensités pour mettre en valeur une région, etc. La modification d'histogrammes est fréquemment utilisée pour effectuer ce type d'opération. En effet, l'histogramme est un outil très utile pour étudier la répartition des composantes d'une image, mais il permet également de corriger le contraste et l'échelle des couleurs pour des images surexposées ou sous-exposées. Sa modification n'altère pas les informations contenues dans l'image, mais elle les rend plus ou moins visibles. On peut donc modifier et améliorer les images si par exemple le contraste est mauvais ou excessif, s'il y a du bruit ou des détails difficilement visibles. La modification d'un histogramme est généralement représentée sur une courbe dite tonale indiquant la modification globale des composantes de l'image avec en abscisse les valeurs initiales et en ordonnées les valeurs après modification. La courbe tonale correspond à une fonction de transfert définie par une table de transcodage appelé look up table, notée LUT. Ainsi, la diagonale indique la courbe telle que les valeurs initiales sont égales aux valeurs finales, c'est-à-dire lorsqu'aucune modification n'a eu lieu.

On peut donc modifier l'aspect d'une image en jouant simplement sur cette courbe afin de faire ressortir certains détails : l'éclaircissement et l'assombrissement ainsi que le contraste.

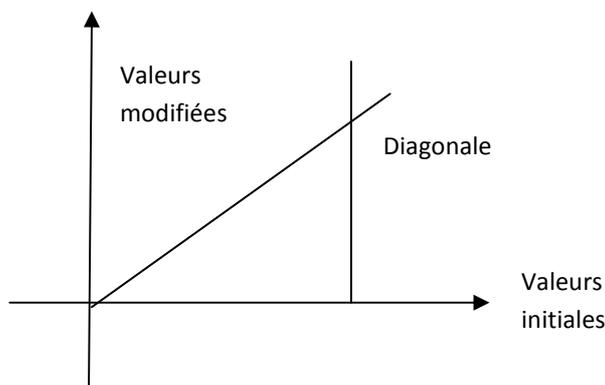


Fig.3 Courbe tonale

3.2 La restauration d'images

Notons que la restauration a pour but de retrouver l'image d'origine à partir de sa version dégradée par ajout d'information au niveau signal connaissant la fonction de dégradation (ou au moins une estimation) et ce, en utilisant une connaissance a priori sur le processus de dégradation. Il s'agit donc, dans un premier temps, de modéliser le processus de dégradation et d'appliquer l'inverse pour retrouver l'image d'origine.

3.3 La segmentation d'images

La segmentation d'images permet de partitionner l'image en régions homogènes de sorte que les régions soient distinctes les unes des autres. Elle peut être définie comme une opération de traitement d'images qui a pour but de séparer des objets du fond ou, d'une manière générale, de rassembler les différents pixels ayant des propriétés communes. La binarisation est une segmentation qui sépare les pixels en deux classes. La première condition demande simplement que la segmentation soit complète : tout pixel de l'image doit appartenir à une région et une seule. Tout algorithme ne doit pas s'arrêter avant d'avoir traité tous les points de l'image. La seconde condition demande que les régions soient des ensembles de points connexes. C'est la raison pour laquelle les algorithmes prennent généralement en compte le voisinage des points. Les critères qui définissent la segmentation sont introduits à la troisième condition. La quatrième condition exprime la maximalité de chaque région.

Plusieurs méthodes de segmentation existent selon les diversités des propriétés recherchées, les problèmes rencontrés et la manière d'opérer le regroupement des pixels. On peut notamment citer :

- La segmentation fondée sur les régions (en anglais : *region-based segmentation*).
- La segmentation fondée sur les contours (en anglais : *edge-based segmentation*)
- La segmentation fondée sur classification ou le seuillage des pixels en fonction de leur intensité (en anglais : *classification* ou *thresholding*)

3.3.1 Segmentation par approche régions

Une région est un ensemble connexe de pixels voisin de l'image ayant des propriétés communes (intensité, texture, etc.) qui différencient des pixels des régions voisines. Les méthodes appartenant à cette approche de segmentation manipulent directement des régions. Les principales méthodes sont les méthodes de type décomposition\fusion, méthodes par croissance de régions et les méthodes fondées sur la modélisation

Parmi les méthodes de type décomposition\fusion on peut citer la méthode *split and merge* proposée par Horowitz et Pavlidis en 1974. Le processus est décomposé en deux étapes. Dans la première étape, ou division, on analyse individuellement chaque région. Si celle-ci ne vérifie pas le critère d'homogénéité, alors on divise cette région en blocs (le plus généralement en 4 quadrants) et l'on réitère le processus sur chaque sous-région prise individuellement.. On peut tout à fait initier le processus en considérant que la première région est composée de toute l'image. Dans la deuxième étape, ou réunion, on étudie tous les couples de régions voisines. Si l'union de ces deux régions vérifie le critère d'homogénéité, alors, on fusionne les régions. La principale difficulté de cette approche réside dans le parcours de l'ensemble de tous les couples de régions voisines. Ce type de méthodes fait souvent appel à la théorie des graphes. On y retrouve les exemples des arbres quaternaires et des approches pyramidales. L'arbre quaternaire (Quad-Tree) est la plus simple implémentation pour la méthode de décomposition de région. C'est un arbre dans lequel chaque nœud représente un bloc. Chaque nœud possède donc 0 sous-nœud (bloc homogène) ou 4 sous-nœuds (bloc non homogène).

En ce qui concerne les méthodes par croissance de régions, ce sont les méthodes de segmentation les plus simples. Elles ont été l'une des premières familles d'outils utilisés pour la segmentation d'image. Dans [1], Zucker introduit les premiers travaux relatifs à la croissance de régions. Cette technique consiste à faire progressivement grossir les régions autour de leur point de départ, son algorithme est composé de deux étapes principales : le choix des points de départ et la croissance de régions. Pour le choix des points de départ appelé aussi « germes », c'est la première et la plus importante étape dans ce genre d'algorithme. L'étape de croissance utilise une mesure de similarité pour choisir les pixels agglomérés. Le choix des points de départ dans les zones les plus homogènes possibles conviendrait mieux pour l'obtention de meilleurs résultats. Si le point de départ est situé dans une zone non homogène, la mesure de similarité va produire de fortes

variations et la croissance va s'arrêter très tôt. Pour retrouver les zones homogènes, on peut réutiliser le principe de décomposition utilisé dans l'algorithme Split/Merge. Comme la décomposition finale nous donne une liste de blocs homogènes, il suffit de choisir le centre des plus gros blocs pour avoir de bons points de départ.

En ce qui concerne l'accroissement ou le grossissement de régions, elle se fait par agglomérations des pixels voisins. Les pixels sont choisis afin de maintenir l'homogénéité de la région. Pour cela, nous devons définir un indicateur d'homogénéité. Les pixels voisins sont ajoutés à la région si l'indicateur d'homogénéité reste vrai. La croissance s'arrête lorsqu'on ne peut plus ajouter de pixels sans briser l'homogénéité.

3.3.2 Segmentation par approche contour

La segmentation par approche contours [2][3] s'intéresse aux contours de l'objet dans l'image. La plupart des algorithmes qui lui sont associés sont locaux, c'est-à-dire qu'ils fonctionnent au niveau du pixel. Les méthodes les plus anciennes utilisent des opérateurs de traitement d'images, tels que le filtre de Canny, pour mettre en évidence les pixels qui semblent appartenir à un contour.

Dans la segmentation par approches contours, il y a deux problématiques à résoudre à savoir la caractérisation de la frontière entre les régions et fermeture des contours. La caractérisation de la frontière consiste à détecter les points contour en utilisant plusieurs méthodes comme les techniques dérivatives, surfaciques et les techniques variationnelles. Quant à la fermeture des contours, il s'agit de lier les points contours non connectés obtenus lors de la phase de détection pour obtenir des contours fermés pour les régions.

3.3.3 Segmentation par classification

Dans cette approche, la segmentation sera vue comme étant un problème de classification de données. La Classification consiste à attribuer à chaque pixel dans l'image une classe (étiquette). Cette affectation peut être effectuée on se basant sur des régions dont on connaît les classes d'appartenance a priori ; on parle de classification supervisée. Dans le cas contraire, on parle de classification non-supervisée ou clustering.

4. Règle de décision de Bayes et réseaux bayésiens

En classification automatique, la classification d'images intervient comme une étape très importante dans des applications de biométrie telles la détection de la peau, la reconnaissance d'empruntes digitales, l'indexation d'image, etc. Elle s'agit d'obtenir des classes dont les éléments (pixels) ont en commun diverses propriétés (intensité, couleur, texture, forme, etc.) telles que deux pixels de la même classe se ressemblent le plus possible et deux pixels de deux classes distinctes se ressemblent le moins possible. Parmi les

méthodes de classification automatique, on peut citer la méthode des K-plus proches voisins (K-PPV), les arbres de décision et la méthode des *k-menas*, etc. On peut aussi se baser sur l'utilisation des réseaux bayésiens et la règle de décision de Bayés

4.1 Décision bayésienne

En effet, une méthode simple de classification supervisée est basée sur le théorème de Bayes. Cette méthode de nature statistique suppose que l'échantillon à analyser est composé de N observations tirées, indépendamment de k classes C_r où $r = 1, 2, \dots, K$ selon les lois de probabilités conditionnelles $p(X/C_r)$ connues. Par application du théorème de Bayes, il est possible de déterminer la probabilité qu'une observation X appartienne à la classe C_r . Cette probabilité, appelée aussi probabilité a posteriori de la classe C_r , est donnée par la relation:

$$P(C_r / X) = \frac{P(x/c_r)p(c_r)}{p(x_r)} \quad (2)$$

Ainsi, une observation X est affectée à la classe C_j si elle vérifie la règle de décision bayésienne suivante :

$$P(X/C_j) P(C_j) \geq P(X/C_r) P(C_r) \quad r \neq j, 1 \leq r \leq k \quad (3)$$

La règle de décision est optimale au sens qu'elle garantit le taux d'erreur de classification minimale par maximisation de la probabilité a posteriori.

4.2 Réseaux bayésiens

Basés sur la décision bayésienne, les réseaux bayésiens adoptent une approche différente par rapport à d'autres méthodes de classification puisqu'ils fournissent des informations utiles sur le problème lui-même. Dans cette approche, on approxime la distribution jointe de probabilités de la classe $P(C, X_1, X_2, \dots, X_k)$ et des attributs. Ici C est la variable aléatoire décrivant la classe et X_1, \dots, X_k les variables aléatoires décrivant les attributs. Ainsi, l'apprentissage avec la classification bayésienne se résume d'abord à l'estimation de cette distribution jointe de probabilité. Après la construction d'une telle estimation, on classe les nouvelles instances en examinant la probabilité conditionnelle de C sachant les valeurs particulières des attributs pour obtenir la classe la plus probable. L'approche standard de la classification bayésienne utilise la règle de chaîne pour décomposer cette distribution jointe de probabilité : $P(C, X_1, X_2, \dots, X_k) = P(C) P(X_1, \dots, X_k / C)$ où $P(C)$ représente la probabilité antérieure (ou à priori) de la classe et $P(X_1, \dots, X_k / C)$, la distribution des valeurs d'attributs sachant la variable classe. Une fois l'estimation de $P(C)$ et de $P(X_1, \dots, X_k / C)$ obtenue, nous pouvons utiliser la règle de Bayes pour calculer la probabilité conditionnelle de la classe sachant les valeurs des attributs. Ainsi, nous pouvons calculer les nouvelles

instances en combinant la probabilité à priori de chaque classe avec les probabilités des valeurs des attributs sachant la valeur de cette classe [4].

5. La couleur en traitement d'images

La couleur est l'un des descripteurs locaux les plus intéressants utilisés en reconnaissance d'objets colorés. Pour caractériser la couleur, la Commission Internationale de l'Eclairage (CIE), organisation internationale, a développée des bases de représentation de la colorimétrie. La colorimétrie standard de la CIE est basée sur le système chromatique à trois composantes rouge, vert et bleu. La fonction d'appariement de couleur est utilisée pour représenter les valeurs de points en fonction des primaires R, V et B calculées telles que :

$$C(\lambda) = R(\lambda)(R) + G(\lambda)(G) + B(\lambda)(B) \quad (4)$$

où :

$$\begin{cases} R = \int_{\lambda=380}^{780} R(\lambda)S(\lambda)d\lambda, \\ G = \int_{\lambda=380}^{780} G(\lambda)S(\lambda)d\lambda, \\ B = \int_{\lambda=380}^{780} B(\lambda)S(\lambda)d\lambda. \end{cases} \quad (5)$$

Dans l'espace **RVB** (Rouge-Bleu- Vert), sont définies les intensités des trois couleurs principales rouge, vert et bleu. Le choix des primaires n'est pas unique. Il est possible de réaliser un changement de primaires soit par une simple transformation linéaire utilisant une matrice de passage ou par une transformation non linéaire. L'espace RVB reste le plus utilisé et le plus simple à manipuler, mais un des principaux inconvénients de cet espace est que certaines couleurs peuvent avoir une composante négative et sont donc non représentables en synthèse additive (fig.4) en plus du fait que les valeurs des composantes trichromatiques liées à la luminance sont corrélées et qu'il existe une multitude de systèmes RGB.

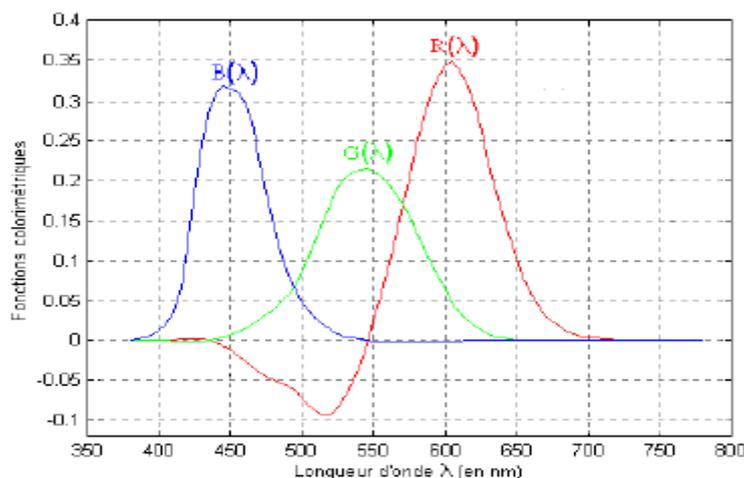


Fig. 4 Les courbes d'appariement $R(\lambda)$, $V(\lambda)$ et $B(\lambda)$

Afin d'obtenir des composantes qui ne tiennent compte que de la chrominance, il convient de normaliser les valeurs des composantes trichromatiques par rapport à la luminance. Ceci est effectué en divisant chaque composante trichromatique par la somme des trois. Les composantes obtenues sont appelées composantes normalisées, forment **l'espace rvg**. Elles sont définies par la relation suivante :

$$\begin{cases} r = \frac{R}{R+G+B}, \\ g = \frac{G}{R+G+B}, \\ b = \frac{B}{R+G+B}. \end{cases} \quad (6)$$

Comme $r+g+b=1$, deux composantes suffisent pour représenter la chrominance d'une couleur.

Par ailleurs, d'autres systèmes de représentation de la couleur ont été conçus sans définir de nouveaux primaires, car ils sont fondés sur le respect de propriétés de notre perception des couleurs. Il existe un très grand nombre de systèmes de représentation de la couleur, chacun possédant ses avantages et ses inconvénients. Ainsi, **le système XYZ** a été défini pour palier certains inconvénients de l'espace RGB, il est constitué de trois primaires X, Y et Z dites virtuelles s'obtient à l'aide d'une matrice de passage A à partir du système RGB en utilisant la relation suivante :

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = A * \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (7)$$

Les coefficients de la matrice A sont directement liés au blanc de référence du système choisi. Les primaires X, Y et Z ont été créés de telle sorte que toutes les couleurs soient exprimées par des composantes trichromatiques positives et de telle sorte que l'une de ces primaires représente une information de luminance. De même que pour le système RGB, la CIE a défini les composantes trichromatiques normalisées (x, y, z) qui se calculent à partir des composantes X, Y et Z selon la relation (8) suivante :

$$\begin{cases} x = \frac{X}{X+Y+Z} \\ y = \frac{Y}{X+Y+Z} \\ z = \frac{Z}{X+Y+Z} \end{cases} \quad (8)$$

Comme $x+y+z=1$, la couleur peut être représentée dans un plan (x, y) qui représente le diagramme de chromaticité

Dans l'espace $\mathbf{L^*u^*v^*}$, le passage des coordonnées x y z vers les coordonnées L^* u^* v^* s'effectue en utilisant les expressions suivantes:

$$\left\{ \begin{array}{l} L^* = \begin{cases} 116 * \left(\frac{Y}{Y_0} \right)^{\frac{1}{3}} - 16 & \text{si } \frac{Y}{Y_0} > 0.008856, \\ 903.3 * \frac{Y}{Y_0} & \text{si } \frac{Y}{Y_0} \leq 0.008856. \end{cases} \\ u^* = 13L^*(u' - u'_0), \\ v^* = 13L^*(v' - v'_0). \end{array} \right. \quad (9)$$

où les quantités u' , v' , v'_0 sont calculées comme suit :

$$\left\{ \begin{array}{l} u' = \frac{4X}{X + 15Y + 3Z}, \\ v' = \frac{9Y}{X + 15Y + 3Z}, \\ u'_0 = \frac{4X_0}{X_0 + 15Y_0 + 3Z_0}, \\ v'_0 = \frac{9Y_0}{X_0 + 15Y_0 + 3Z_0}. \end{array} \right. \quad (10)$$

où les trois composantes (X_0, Y_0, Z_0) représentent le blanc de référence pris dans l'espace XYZ.

Pour sa part, l'espace $\mathbf{L^*a^*b^*}$ est établi par des relations non linéaires à partir du système XYZ. L'information luminance est calculée de la même façon que le système $L^*u^*v^*$ et les composantes chrominances pour ce système sont données par la relation suivante :

$$\left\{ \begin{array}{l} a^* = 500 \left[f\left(\frac{X}{X_0}\right) - f\left(\frac{Y}{Y_0}\right) \right], \\ b^* = 300 \left[f\left(\frac{Y}{Y_0}\right) - f\left(\frac{Z}{Z_0}\right) \right]. \end{array} \right. \quad (11)$$

avec :

$$f(x) = \begin{cases} x^{1/3} & \text{si } x > 0.008856 \\ 7.787x + \frac{16}{116} & \text{si } x \leq 0.00856 \end{cases} \quad (12)$$

Pour ces deux systèmes uniformes, la composante luminance représente la clarté. La première composante de chrominance (u^* et a^*) représente

l'opposition vert-rouge alors que la seconde composante (v^* et b^*) représente une opposition de couleur bleu-jaune.

Le **système ITS** modèle correspond à une expression des composantes I, T et S dans le cube de couleurs d'un système RGB. Sa modélisation est issue de la déformation du cube en le renversant sur le coin représentant le noir. L'axe achromatique dans le cube des couleurs correspond à l'axe d'intensité dans le système ITS. Soient un point P dans le cube des couleurs RGB, O' la projection orthogonale de P sur l'axe achromatique. L'intensité I correspond à la norme du vecteur OO' (O l'origine du cube $R=G=B=0$). La saturation correspond aux rapports entre la norme du vecteur OP, distance entre le point P et l'axe achromatique et la norme du vecteur OO' qui représente l'intensité s. La teinte correspond à l'orientation du vecteur OP (fig.5).

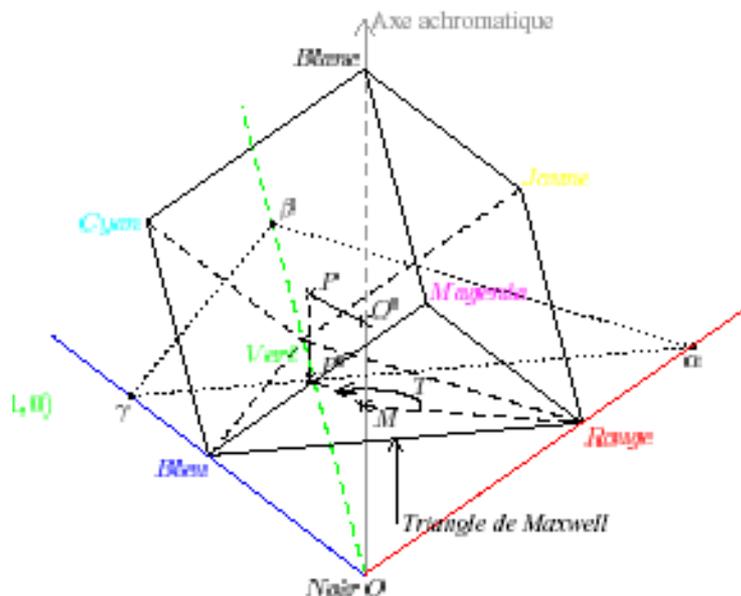


Fig.5 Cube des couleurs dans le modèle triangulaire.

Les formules exprimant les composantes I, T et S ou I, H et S en anglais, sont :

$$\left\{ \begin{array}{l} I = \frac{R+G+B}{3}, \\ S = 1 - \frac{3 * \min(R, G, B)}{R+G+B}, \\ H = \arccos \left(\frac{0.5 * (R-G) + (R-B)}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right). \end{array} \right. \quad (13)$$

En projetant les points du cube des couleurs suivant l'axe achromatique sur le plan perpendiculaire à cet axe et passant par le point blanc, nous obtenons une surface fermée hexagonale dont le centre est le point blanc (fig.6). Ce système a été défini par Travis. Il représente la couleur sous forme d'un triplet teinte H (hue), saturation S et luminosité V (value) pour former **l'espace HSV**. Pour un point couleur RGB, sa projection orthogonale sur l'axe achromatique donne son intensité.

La teinte est définie comme l'angle entre les vecteurs $\overrightarrow{\text{Blanc M}}$ et $\overrightarrow{\text{Blanc Rouge}}$.

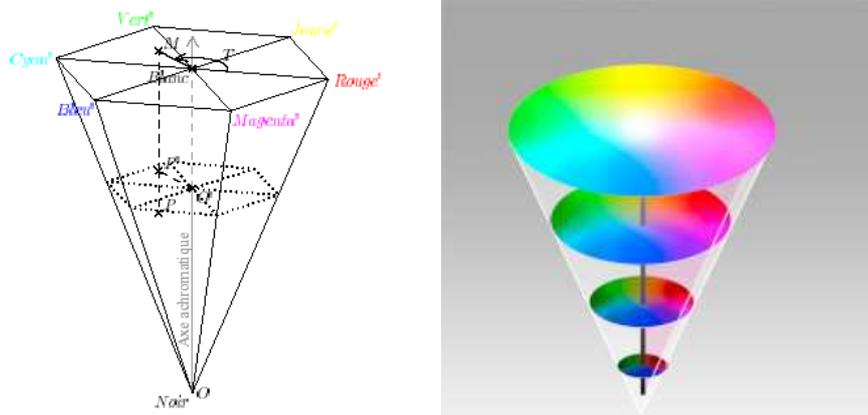


Fig.6 Modèle du cône hexagonal.

Dans ce modèle les composantes V, S et H sont données par la relation suivante :

$$\left\{ \begin{array}{l} V = \max(R, G, B), \\ S = \frac{V - \min(R, G, B)}{V} \\ H = \begin{cases} \frac{G - B}{V - \min(R, G, B)} & \text{si } V = R, \\ 2 + \frac{B - R}{V - \min(R, G, B)} & \text{si } V = G, \\ 4 + \frac{R - G}{V - \min(R, G, B)} & \text{si } V = B. \end{cases} \end{array} \right. \quad (14)$$

Une autre possibilité intéressante, dans l'optique d'un traitement de la couleur, est de décorréler la chrominance de la luminance. Ainsi l'information chrominance est portée sur deux axes, et l'information luminance sur la troisième. Nous pouvons ici citer différents espaces comme YCbCr, YIQ, YUV, ou encore AC1C2. YCbCr est obtenu via la formule suivante :

$$\begin{pmatrix} Y \\ Cr \\ Cb \end{pmatrix} = \begin{pmatrix} 0.299 & 0.582 & 0.114 \\ 500 & -0.419 & -0.081 \\ -0.169 & -0.331 & 0.500 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (15)$$

6. Conclusion

Dans ce chapitre, nous avons vu les principales étapes et techniques de traitement d'image. Nous avons particulièrement évoqué les opérations les plus usitées en traitement d'images et tout particulièrement la segmentation qui est une opération parmi les plus importantes. Nous avons ensuite décrit brièvement les espaces de couleur qui sont une notion importante pour traiter des objets de ce type.

Ainsi, nous concluons qu'il n'existe pas de loi pour un bon choix d'une méthode de segmentation ni de choix d'un espace de couleur. Ce choix peut être toutefois justifié. Ainsi, nous consacrons le prochain aux réseaux bayésiens qui constituent un choix pour une application à la reconnaissance de la peau.

Chapitre 2

Généralités sur les réseaux bayésiens

1. Introduction

Les réseaux bayésiens sont initialement connus comme « Réseaux de Relation » mais sont nommés « Bayes » après la mort du théologien et mathématicien le révérent Thomas Bayes (1702-1761) qui a écrit une loi de base de probabilité maintenant appelée « théorème de Bayes », présenté à titre posthume en 1763. Les réseaux bayésiens sont le résultat de recherches effectuées dans les années 1980, dues à J. Pearl à UCLA (University of California, Los Angeles) et à une équipe de recherche danoise à l'université d'Aalborg. L'objectif initial de ces travaux était d'intégrer la notion d'incertitude dans les systèmes experts. Les chercheurs se sont rapidement aperçus que la construction d'un système expert nécessitait presque toujours la prise en compte de l'incertitude dans le raisonnement. En effet, ces modèles sont de plus en plus utilisés pour représenter l'incertitude des connaissances dans des systèmes complexes. Ils ont émergé de l'association entre la théorie des probabilités et la théorie des graphes pour donner un outil graphique de représentation des probabilités jointes sur un ensemble de variables aléatoires. Un réseau bayésien est un outil de modélisation qui peut être continu ou discret et les modalités entre ses variables peuvent être d'ordre causal, temporel, hiérarchique,..., etc. En général, une seule modalité est utilisée dans un même réseau et dans la plus part du temps, il s'agit de la causalité. Parfois les réseaux bayésiens ne sont pas utilisés pour représenter ces modalités, mais juste pour profiter de leur avantage calculatoire.

Dans ce présent chapitre, nous exposons la représentation graphique de la causalité et la représentation probabiliste associée. Quelques lois de probabilités seront aussi exposées. En suite, nous nous attardons sur des généralités sur les réseaux bayésiens tout en s'intéressant aux algorithmes d'inférence et d'apprentissage et cela, avant d'entamer les différentes extensions de ces réseaux.

2. Représentation graphique de la causalité

La représentation graphique la plus intuitive de l'influence d'un événement, d'un fait, ou d'une variable sur une autre, est probablement de relier la cause à l'effet par une flèche orientée (fig.1).

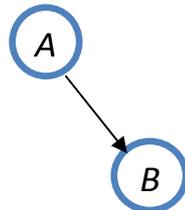


Fig.1 Représentation graphique de la causalité

Dans ce schéma, A et B se sont des événements, qui peuvent être observés ou non, vrais ou faux. Du point de vue du sens commun, le graphe ci-dessus peut se lire comme ceci. «La connaissance de A détermine la connaissance de B». Bien que la flèche soit orientée de A vers B, elle peut cependant fonctionner dans les deux sens. En d'autres termes, toute information sur A peut modifier la connaissance sur B, et réciproquement, toute information sur B peut modifier la connaissance sur A.

2.1 Circulation de l'information dans un graphe causal

La représentation en graphe est un moyen de renseigner sur les sens de circulation de l'information entre les différentes variables. Pour étudier de plus près comment l'information circule au sein d'un graphe causal, nous allons utiliser un exemple, extrêmement classique dans la littérature sur les réseaux bayésiens proposés par J. Pearl dans [5] :

« Ce matin-là, alors que le temps est clair et sec, M.Holmes sort de sa maison. Il s'aperçoit que la pelouse de son jardin est humide. Il se demande alors s'il a plu pendant la nuit ou s'il a simplement oublié de débrancher son arroseur automatique. Il jette alors un coup d'œil à la pelouse de son voisin, M. Watson, et s'aperçoit qu'elle est également humide. Il en déduit alors qu'il a probablement plu, et il décide de partir au travail sans vérifier son arroseur automatique. »

La représentation graphique du modèle causal utilisé par M. Holmes est la suivante (fig.2) :

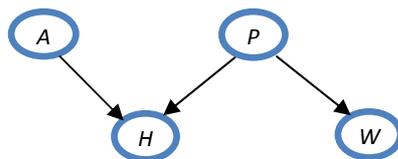


Fig.2 Modélisation de l'exemple de la pelouse mouillée [5]

A	M. Holmes a oublié de débrancher son arroseur automatique
P	Il a plu pendant cette nuit
H	L'herbe du jardin de Mr Holmes est humide
W	L'herbe du jardin de Mr Watson est humide

Tab 1 les variables de l'exemple de la pelouse mouillée

Nous savons que H est vrai. Le modèle nous indique que H a dû être causé soit par A, soit par P. faute d'information complémentaire, les deux causes sont a priori également plausibles. Le fait que W soit également vrai renforce la croyance en P et réfute la croyance en A. Dans cet exemple simple, on voit que l'information a circulé uniquement dans le sens effet→cause. Donc il est essentiel de noter que l'information ne circule pas seulement dans la direction de l'orientation des arcs.

2.2 Différents types de connexion

Dans l'exemple ci-dessus, nous avons vu qu'une information certaine se propage dans un graphe en modifiant les croyances que nous avons des autres faits. Pour des graphes plus complexes, la circulation de l'information dépend aussi de notre connaissance sur les variables du problème. Le graphe peut en effet contenir des variables dont nous connaissons les valeurs, des variables observées et des variables dont nous ne connaissons pas les valeurs des variables cachées. Nous allons étudier quels chemins cette information peut prendre à l'intérieur d'un graphe. Nous allons considérer les trois cas suivants, qui décrivent l'ensemble des situations possibles faisant intervenir trois événements.

- *Connexion convergente*

L'information ne peut circuler de X à Y que si Z est connu.

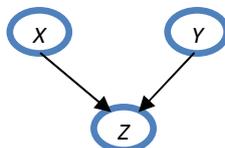


Fig.3 Connexion convergente

- *Connexion en série*

L'information ne peut circuler de X à Y que si Z n'est pas connu.

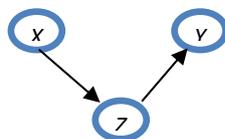


Fig.4 Connexion en série

- *Connexion divergente (V-structure)*

L'information ne peut circuler de X à Y que si Z n'est pas connu.

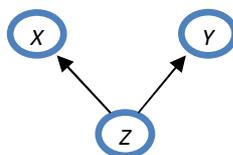


Fig.5 Connexion divergente

2.3 Notion de d-séparation

Dans un graphe relativement complexe pour lequel nous disposons d'un certain nombre d'informations, nous ne savons pas si une nouvelle information rapportera un changement sur l'ensemble des autres nœuds de ce graphe. Pour cela, nous définissons la règle de d-séparation (directed-separation) qui décrit dans quelles conditions l'information peut circuler entre un nœud père et un nœud fils.

Définition 1 [6]

Soient (X, Y, Z) des nœuds d'un graphe G , on dit que X et Y soit d-séparer par Z si, pour tout chemin entre X et Y , l'une au moins des deux conditions suivantes est vérifiée :

- Le chemin converge en un nœud W , tel que $W \neq Z$, et W n'est pas une cause directe de Z
 - Le chemin passe par Z , et est soit divergent, soit en série au nœud Z .
- « X est d-séparé de Y par Z » se note par « $X|Y|Z$ »
Si X et Y ne sont pas d-séparés, on dit qu'ils sont d-connectés.

Dans l'exemple de la pelouse mouillée, si P est la seule information connue, la connaissance sur W ne modifiera en rien la connaissance sur H : le circuit d'information de W à H est bloqué par P , on peut dire alors que W et H sont d-séparés par P ($W|P|H$).

3. Représentation probabiliste associée

Avec la représentation graphique de la causalité, on peut connaître la direction de circulation de l'information dans le graphe, mais on ne peut pas connaître la quantité de cette circulation d'information. Alors, il faut une représentation probabiliste associée à ce graphe et cela en assimilant ce que nous appelons la croyance en un fait ou la plausibilité d'un fait à une probabilité mathématique. Avec la relation cause \rightarrow effet : $A \Rightarrow B$ et en utilisant les algorithmes d'inférence, on peut représenter la quantité de circulation de l'information par la probabilité conditionnelle $P(B/A)$ tout en respectant les règles de transposition des graphes en termes de probabilités.

3.1 Indépendance conditionnelle

L'indépendance conditionnelle est un concept qui permet de simplifier le calcul d'une loi jointe en décomposant la probabilité jointe dépendant de toutes les variables en un produit de probabilités qui ne dépendent que d'un nombre réduit de variables. On parle d'indépendance conditionnelle lorsque l'occurrence d'un événement ne dépend pas d'un autre événement.

Définition.2 [7] : Si (Ω, Z, P) est un espace probabiliste, soient A, B, C trois éléments de Z . On dit que A et B sont indépendants conditionnellement à C et on note $A \perp B / C$ si et seulement si l'une des propriétés suivantes est vérifiée [7]:

- $P(A/B, C) = P(A/C)$ (connaissant C , la connaissance de B n'apporte rien à la connaissance de A)

$$\blacksquare P(A, B/C) = P(A/C) P(B/C)$$

3.2 Théorème de Bayes

Nous avons la probabilité conditionnelle

$$P(A/B) = \frac{P(A,B)}{P(B)} ; \quad P(A) \neq 0, P(B) \neq 0 \quad (1)$$

Comme $P(A, B) = P(B, A)$, on en tire immédiatement:

$$P(A/B) P(B) = P(B/A) P(A) \quad (2)$$

d'où la loi de Bayes [20]:

$$P(A/B) = \frac{P(B/A) P(A)}{P(B)} \quad (3)$$

$P(A/B)$ est la probabilité a posteriori de A sachant B, Elle est « postérieure », au sens qu'elle dépend directement de B. Le terme $P(A)$ est la probabilité a priori de A. Elle est « antérieure » au sens qu'elle précède toute information sur B. Elle est aussi appelée la probabilité marginale de A. Le terme $P(B/A)$, pour un B connu est appelée la fonction de vraisemblance de A. Le terme $P(B)$ est appelé la probabilité marginale ou a priori de B.

La loi de Bayes est souvent utilisée dans l'estimation de modèle à partir de données observées. Ainsi, en disposant d'un ensemble de données D , nous cherchons à estimer le modèle M qui représente les données. Dans ce cas, la loi de Bayes s'exprime sous la forme suivante :

$$P(M/D) = \frac{P(D/M) P(M)}{P(D)} \quad (4)$$

La connaissance à priori sur les données $P(D)$ peut se décomposer comme la somme des connaissances sur les données par rapport à tous les modèles si les variables sont continues ou comme une intégrale si les variables sont discrètes. Cette opération s'appelle la *marginalisation* [8] :

$$P(D) = \sum_M P(D/M) P(M) \quad (5)$$

$$P(D) = \int_M P(D/M) P(M)$$

Théorème de Bayes généralisé (règle de chaînage) : La distribution de probabilités jointes globales (associée à l'ensemble des variables aléatoires) est donnée selon le théorème de Bayes par :

$$\begin{aligned} P(x_1, \dots, x_n) &= P(x_1) P(x_2/x_1) P(x_3/x_1, x_2) \dots p(x_n/x_1, x_2, \dots, x_{n-1}) \\ &= P(x_1) \prod_{i=2}^n P(x_i/x_{i-1}, \dots, x_1) \end{aligned} \quad (6)$$

Mais dans un réseau bayésien, les probabilités conditionnelles de certaines variables x_i ne sont pas dépendantes de tous les prédécesseurs de x_i (càd x_1, \dots, x_{i-1}) mais seulement de certains d'entre eux. En d'autres termes, supposons que x_i soit indépendante de tous ses prédécesseurs sauf de certains d'entre eux : ce qui ont une influence directe sur x_i . Nous appellerons cet ensemble restreint les parents markoviens de x_i (où simplement les parents) $Pa(x_i)$, alors nous pouvons écrire :

$$P(x_i/x_1 \dots x_n) = P(x_i/Pa(x_i)) \quad (7)$$

D'où la loi jointe est décomposée sous forme de produit des distributions de probabilités locales par la règle de chaînage :

$$P(x_1, \dots, x_n) = \prod P(x_i / \text{parents}(x_i)) \quad (8)$$

(Cette formule est également appelée formule de factorisation ou de chaînage).

4 Réseaux bayésiens

Un réseau bayésien est constitué de :

- Un graphe orienté, acyclique dont les nœuds sont un ensemble de variables aléatoires $X = x_1, x_2, \dots, x_n$, les arcs sont des relations de dépendance entre ces variables. L'ensemble des nœuds et des arcs forme ce que l'on appelle la structure du réseau bayésien. C'est la représentation qualitative de la connaissance.
- Un ensemble de distribution de probabilité. Pour chaque nœud, on dispose d'une table de probabilité conditionnelle P (variable/parents (variable)). C'est la représentation quantitative de la connaissance.

Les deux représentations qualitative et quantitative de la connaissance forment les paramètres qualitatifs et quantitatifs respectivement du réseau bayésien.

Théorème 1 [8]

Soient :

- Un graphe acyclique orienté $G = (V, A)$
- Une épreuve ε à laquelle est associée un espace probabiliste fini (Ω, Z, P) et n variables aléatoires $(X_i)_{1 \leq i \leq n}$

G et ε définissent un réseau bayésien que l'on note $B = (G, P)$ si et seulement si :

- Il existe une bijection entre les nœuds du graphe G et les variables (X_i)
- La propriété de factorisation est vérifiée : $P(X_1, \dots, X_n) = \prod P(X_i \mid \text{parents}(X_i))$

Pour l'utilisation des réseaux bayésiens deux étapes sont nécessaires, la première étape dite étape d'apprentissage lors de laquelle il y a construction du réseau à partir de connaissances a priori ou en faisant un apprentissage des paramètres du réseau à partir d'une base de données. Une fois que le réseau est construit, vient ensuite l'étape d'utilisation même du réseau. Cette étape s'appelle inférence. Lors de l'inférence, on dispose d'un certain nombre de variables observées et l'on cherche à utiliser ces observations pour déterminer une connaissance sur les variables non observées (fig.6). La transposition d'un graphe acyclique orienté en espace de probabilités conduit à des résultats conformes au raisonnement intuitif que l'on peut mener directement sur ce graphe. Cela va permettre de ramener l'ensemble des problèmes de l'inférence à des problèmes de théorie des graphes. La puissance des réseaux bayésiens apparait pleinement lorsqu'on réalise que la théorie de probabilités permet de propager de manière cohérente l'impact des observations faites sur l'ensemble des issues incertaines. C'est ainsi que les réseaux bayésiens peuvent déjouer les erreurs communes faites dans des raisonnements intuitifs dues à une mauvaise connaissance des probabilités. Cependant, ces problèmes restent relativement complexes, et constituent des sujets de recherche très active. Les calculs mis en œuvre, même sur des cas très simples, sont lourds (NP-difficile). Pour cela, nous étudions dans la section suivante la notion d'inférence qui a pour but de faciliter les calculs et la propagation de l'information dans un réseau bayésien.

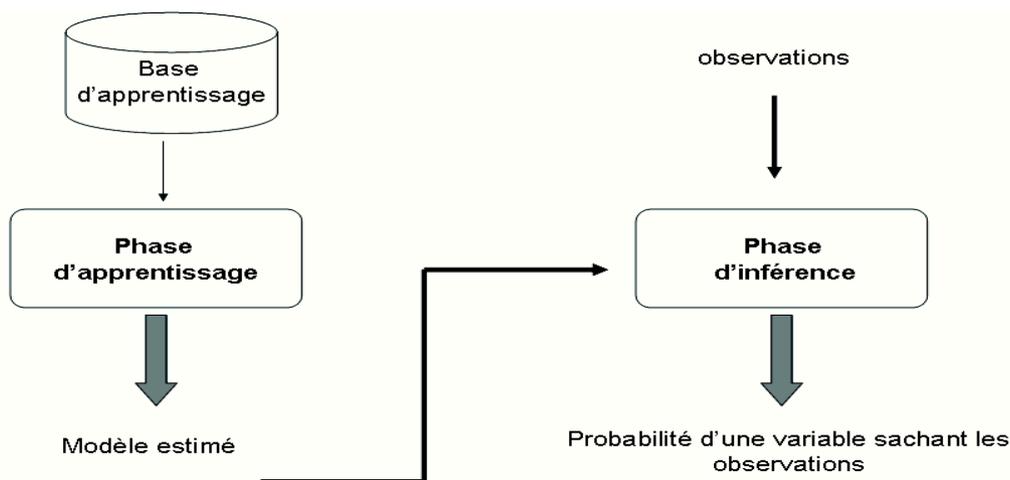


Fig.6 Processus de fonctionnement d'un réseau bayésien

5. L'inférence dans les réseaux bayésiens

Les réseaux bayésiens ont été développés pour tenter de résoudre certains problèmes de prédictions et d'abduction courants en intelligence artificielle (IA). Dans ce type de tâches, il est nécessaire de trouver une interprétation cohérente des observations avec les données connues à priori. Cette cohérence est appelée *inférence*. L'inférence probabiliste signifie donc le calcul de $p(X/Y)$ où X un ensemble de variables et Y un ensemble d'observations décrivant le problème. Ces calculs sont reconsidérés à chaque nouvelle observation, la raison pour laquelle l'inférence est aussi dite « mise à jours des probabilités ». Donc d'une manière générale, tout calcul portant sur le problème de l'inférence est uniquement un problème de calcul. Il n'y a aucun problème théorique. Le calcul d'inférence étant en général NP-difficile, deux types de méthodes ont été développés: il y a les approches complètes et approximatives appelées aussi méthodes exactes et approchées.

5.1 Les méthodes exactes

Les premiers algorithmes ont été proposés dans [5] concernant le cas d'arbre et de polyarbre, en suite généralisés par [9] c'est le cas d'arbre de Jonction (JT). Par ailleurs, ces deux algorithmes et l'algorithme d'élimination de variables (Bucket Elimination) [10] sont les algorithmes d'inférence exacte les plus réputés. Ils seront détaillés dans la section ci-dessous.

5.1.1 Propagation de messages locaux dans un arbre

La méthode de propagation de messages locaux ou de connaissance dans un arbre a été introduite par Kim & Pearl (1983)[11] ; cette méthode est connue sous le nom de « polytree algorithm » elle se base sur la transmission de messages entre nœuds voisins, transitant par les arcs entre ces nœuds. Le but de cette méthode est que chaque nœud communique à ses voisins les informations qu'il a collectées, jusqu'à ce que tout nœud puisse mettre à jour sa probabilité marginale en fonction de l'ensemble de l'information qu'on note \mathcal{E} reçue par le graphe. Cette méthode se

complice nettement lorsque le graphe n'est pas un arbre. Dans ce cas, nous nous restreindrons au cas d'un arbre.

- Soit $G = (V, E)$ un graphe acyclique orienté (un arbre), où V est l'ensemble de noeuds, et E celui des arcs.
- Soit $X = \{X_v : v \in V\}$ un ensemble de variables aléatoires. Chaque variable X_v correspond à un noeud v du graphe. Pour chaque noeud $v \in V$, on définit π_v , l'ensemble de ses parents dans le graphe,
- Soit θ , l'ensemble de probabilités conditionnelles $\{\theta_v\} = \{P(x_v | x_{\pi_v})\}$, $v \in V$, alors le couple (G, θ) définit un réseau Bayésien.
- On peut partitionner l'ensemble de l'information \mathcal{E} sur l'ensemble du graphe en trois sous-ensembles différents relativement à la variable X_v associée au noeud v du graphe $G = (V, E)$:
 - $\mathcal{E}_{X_v}^+$ Correspond aux informations élémentaires sur les noeuds dont la chaîne vers v passe par un parent de v ;
 - $\mathcal{E}_{X_v}^-$ Correspond aux informations élémentaires sur les noeuds dont la chaîne vers v passe par un enfant de v ;
 - $\mathcal{E}_{X_v}^0$ Corresponds aux informations élémentaires sur le noeud v lui-même.

On distingue deux types de messages, λ et π :

$$\lambda(x_v) = P(\mathcal{E}_{X_v}^- \setminus x_v) \quad (9)$$

$$\pi(x_v) = P(x_v \setminus \mathcal{E}_{X_v}^+) \quad (10)$$

$\lambda(x_v)$ et $\pi(x_v)$ sont généralement représentés par des vecteurs, où chaque élément est associé à une valeur de X_v :

$$\lambda(x_v) = [\lambda(X_v=x_v^1), \dots, \lambda(X_v=x_v^n)] \quad (11)$$

$$\pi(x_v) = [\pi(X_v=x_v^1), \dots, \pi(X_v=x_v^n)] \quad (12)$$

Pour résoudre le problème de l'inférence, il faut calculer la probabilité $P(X_v \setminus \mathcal{E})$:

En appliquant le théorème de Bayes :

$$\begin{aligned} P(x_v \setminus \mathcal{E}) &= P(X_v \setminus \mathcal{E}_{X_v}^+, \mathcal{E}_{X_v}^-) = \frac{p(\mathcal{E}_{X_v}^- / X_v, \mathcal{E}_{X_v}^+) p(X_v / \mathcal{E}_{X_v}^+)}{p(\mathcal{E}_{X_v}^-, \mathcal{E}_{X_v}^+)} \\ &= \alpha P(\mathcal{E}_{X_v}^- / X_v, \mathcal{E}_{X_v}^+) P(x_v / \mathcal{E}_{X_v}^+) \end{aligned} \quad (13)$$

tel que :

$$\alpha = \frac{1}{p(\mathcal{E}_{X_v}^-, \mathcal{E}_{X_v}^+)} \quad \text{qui est une constante}$$

Comme $\mathcal{E}_{X_v}^-$ et $\mathcal{E}_{X_v}^+$ sont d-séparés (indépendants conditionnellement) par X_v nous aurons :

$$p(\mathcal{E}_{X_v}^- / X_v, \mathcal{E}_{X_v}^+) = p(\mathcal{E}_{X_v}^- / X_v) \quad (14)$$

et

$$p(\mathcal{E}_{X_v}^- / \mathcal{E}_{X_v}^+) = p(\mathcal{E}_{X_v}^-) \quad (15)$$

d'où

$$P(x_v \setminus \mathcal{E}) = \alpha P(\mathcal{E}_{X_v}^- / X_v) P(x_v / \mathcal{E}_{X_v}^+) = \alpha \lambda(X_v) \pi(X_v) \quad (16)$$

donc

$$P(X_v \setminus \mathcal{E}) \propto \lambda(X_v) \pi(X_v) \quad (17)$$

où \propto signifie la proportion.

Pour calculer $\lambda(X_v)$ et $\pi(X_v)$, X_v doit recevoir des informations de tout le réseau. Pour cela, les étapes suivantes sont nécessaires pour la propagation de l'information :

1. Le nœud X_v demande à ses voisins de lui envoyer leurs messages pour qu'il puisse calculer $P(X_v|\mathcal{E})$,
2. X_v a besoin des λ -messages de tous ses enfants pour pouvoir calculer $\lambda(X_v)$
3. Il a aussi besoin des π -messages de tous ses parents pour pouvoir calculer $\pi(X_v)$
4. Les enfants de X_v , Y_1, \dots, Y_m , qui veulent envoyer les λ -messages ($\lambda_{Y_j}(X_v)$) vers leur parent X_v ont besoin des λ -messages de tous leurs enfants et de π -messages de tous leurs parents sauf X_v
5. Les nœuds U_1, \dots, U_n qui veulent envoyer un π -message ($\pi_{X_v}(U)$) vers leur enfant X_v ont besoin des π -messages de tous leurs parents et des λ -messages de tous leurs enfants sauf X_v
6. Lorsque X_v reçoit tous les messages qu'il avait demandés, il les multiplie avec sa table de probabilité conditionnelle (loi jointe). Puis, il effectue des sommes (loi marginale) pour obtenir la probabilité voulue $P(X_j|\mathcal{E})$.

Calcul de $\lambda(X_v)$

- Si X_v est une variable observée $X_v = x_v^0$ alors $\lambda(X_v)$ est un vecteur de taille égale au domaine de X_v . Ce vecteur vaut 0 partout sauf à la place de la valeur observée où il vaut 1. C'est-à-dire :

$$\begin{aligned}\lambda(X_v) &= 0 \text{ si } X_v \neq x_v^0 \\ &= 1 \text{ si } X_v = x_v^0\end{aligned}\quad (18)$$

- Si le nœud v est une feuille de l'arbre, alors le vecteur $\lambda(X_v)$ vaut 1 partout.
- Sinon, $\lambda(X_v)$ est calculé en utilisant $\lambda(Y_1), \dots, \lambda(Y_m)$ tel que Y_1, \dots, Y_m sont les enfants de X_v .

Soit $\mathcal{E}_{X_v}^- = \bigcup_{1 \leq j \leq m} \mathcal{E}_{Y_j}^-$

Alors :

$$\begin{aligned}\lambda(X_v) &= P(\mathcal{E}_{X_v}^- \setminus X_v) \\ &= P(\mathcal{E}_{Y_1}^- \dots, \mathcal{E}_{Y_m}^- \setminus X_v) \\ &= P(\mathcal{E}_{Y_1}^- \setminus X_v) \dots P(\mathcal{E}_{Y_m}^- \setminus X_v) \\ &= \lambda_{Y_1}(X_v) \dots \lambda_{Y_m}(X_v) \\ &= \prod_{1 \leq j \leq m} \lambda_{Y_j}(X_v)\end{aligned}\quad (19)$$

On peut calculer $\lambda_{Y_j}(X_v)$ de la manière suivante :

$$\begin{aligned}\lambda_{Y_j}(X_v) &= P(\mathcal{E}_{Y_j}^- \setminus X_v) \\ &= \sum_{Y_j} P(\mathcal{E}_{Y_j}^-, Y_j \setminus X_v) \\ &= \sum_{Y_j} P(\mathcal{E}_{Y_j}^- \setminus Y_j, X_v) P(Y_j \setminus X_v) \\ &= \sum_{Y_j} P(\mathcal{E}_{Y_j}^- \setminus Y_j) P(Y_j \setminus X_v) \\ &= \sum_{Y_j} \lambda(Y_j) P(Y_j \setminus X_v)\end{aligned}\quad (20)$$

On obtient finalement que:

$$\lambda(X_v) = \prod_{1 \leq j \leq m} \sum_{Y_j} \lambda(Y_j) P(Y_j \setminus X_v)\quad (21)$$

Calcul de $\pi(X_v)$

- Si la variable X_v est observée, alors $\pi(X_v) = \lambda(X_v)$ est un vecteur de taille égale au domaine de X_v . Ce vecteur vaut 0 partout sauf à la place de la valeur observée où il vaut 1.
- Si le nœud v est la racine de l'arbre, alors $\pi(X_v) = P(X_v)$
- Sinon, $\pi(X_v)$ se calcul en passant par le calcul de π des parents U_1, \dots, U_n de X_v :

$$\begin{aligned}
 \pi(X_v) &= P(X_v \setminus \mathcal{E}_{X_v}^+) \\
 &= \sum_{U_i} P(X_v, U_i \setminus \mathcal{E}_{X_v}^+); i=0, \dots, n \\
 &= \sum_{U_i} P(X_v \setminus U_i, \mathcal{E}_{X_v}^+) P(U_i \setminus \mathcal{E}_{X_v}^+) \\
 &= \sum_{U_i} P(X_v \setminus U_i) P(U_i \setminus \mathcal{E}_{X_v}^+) \\
 &= \sum_{U_i} P(X_v \setminus U_i) P(U_1 \setminus \mathcal{E}_{X_v}^+) \dots P(U_n \setminus \mathcal{E}_{X_v}^+)
 \end{aligned} \tag{22}$$

et finalement:

$$\pi(X_i) = \sum_{U_i} P(X_v \setminus U_i) \prod_{1 \leq i \leq n} \pi(U_i) \tag{23}$$

La figure 7 représente le mécanisme de propagation de messages dans un réseau à structure d'arbre.

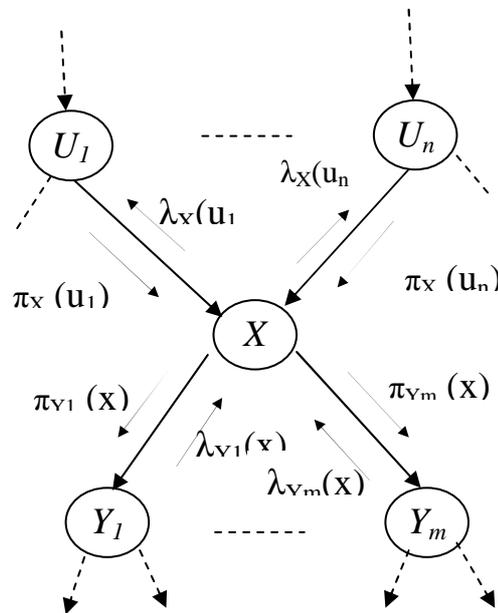


Fig 7 Mécanisme de propagation de connaissances au niveau d'un nœud dans un réseau bayésien à structure d'arbre

5.1.2 Inférence dans une structure générique : Algorithme de Jtree

L'algorithme de propagation de message de Pearl ne s'applique qu'aux arbres, une généralisation a donc été proposée c'est l'algorithme d'arbre de jonction, appelée aussi « clustering ou clique-tree propagation algorithm » ou « méthode JLO ». Elle a été introduite par Lauritzen & Spiegelhalter (1988)[12] et Jensen, Lauritzen & Olsen (1990)[13]. Cette méthode est applicable pour toute structure DAG, son principe est de transformer le graphe d'une structure générique en un arbre non orienté appelé arbre de jonction. Cette transformation opère en trois étapes : la moralisation, la triangulation et la construction de l'arbre de jonction. La moralisation consiste à obtenir un graphe moral en désorientant les arcs et en mariant les parents d'un même nœud. Le graphe moral a donc la propriété de créer une clique pour chaque nœud et ces parents. La triangulation est l'opération qui a

pour but d'obtenir un graphe triangulé à partir d'un graphe non orienté. Cette étape consiste donc à extraire du graphe moralisé G^m un ensemble de cliques de nœuds. La construction de l'arbre de jonction est une étape qui consiste à avoir une structure arborescente ayant l'ensemble des cliques d'un graphe triangulé comme ensemble de nœuds tel que toute clique sur le chemin entre deux cliques C_1 et C_2 doit contenir les nœuds de $C_1 \cap C_2$ (propriété de l'intersection courante).

Après la construction de l'arbre de jonction, l'algorithme de passage de messages est lancé sur cette dernière, il permet de calculer les probabilités marginales de tous les nœuds pour chaque clique. La complexité de cet algorithme est déterminée par la plus grande clique.

Tout comme le graphe triangulé n'est pas unique, l'arbre de jonction ne l'est pas non plus. Par ailleurs, chaque nœud du graphe triangulé doit apparaître dans au moins une clique utilisée dans la construction de l'arbre de jonction.

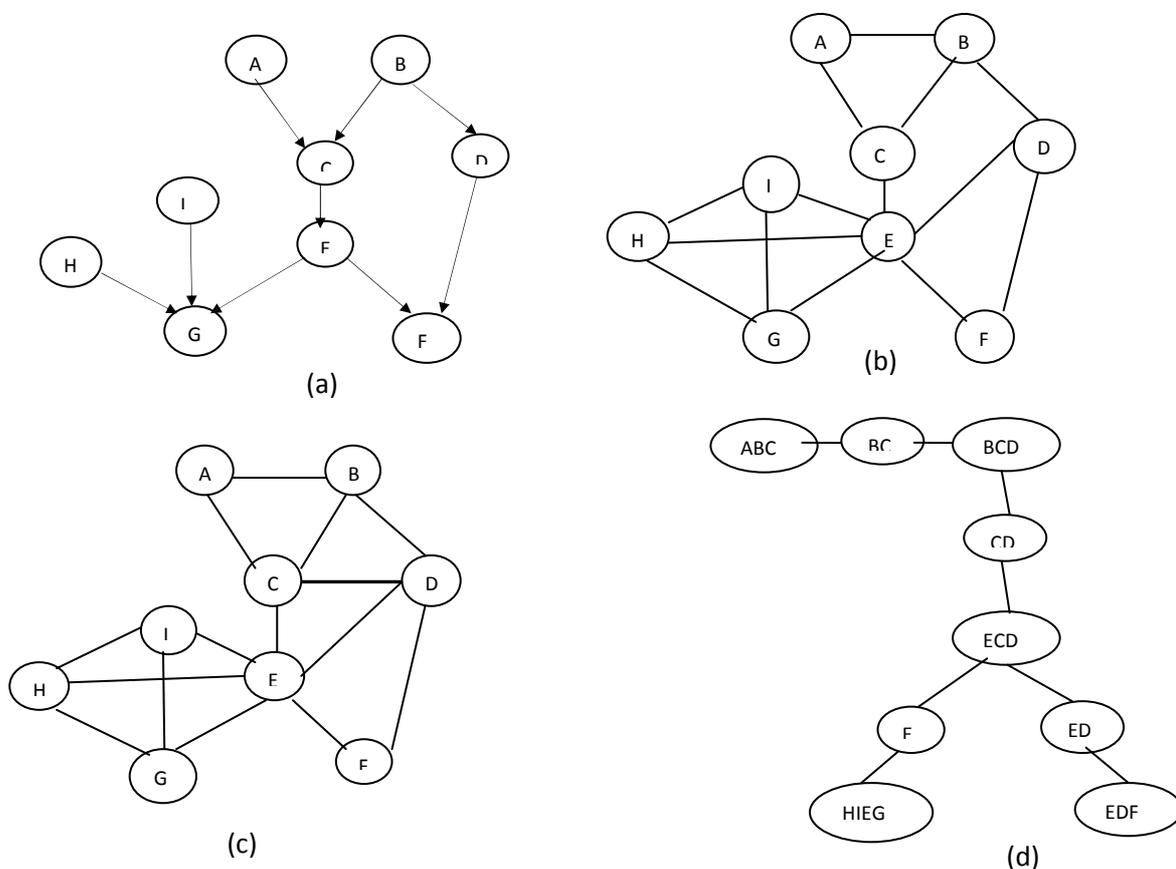


Fig.8 Etape de l'algorithme Jtree. (a) graphe original, (b) graphe moralisé, (c) graphe triangulé, (d) arbre de jonction

5.1.3 Conditionnement global

L'algorithme de propagation de messages locaux ne s'applique que sur les réseaux bayésiens dont le graphe est un arbre. Le problème se pose de l'étendre à tout réseau bayésien. Le conditionnement global est une méthode également proposée par [5] dont le principe est de changer la connectivité du réseau en trouvant un sous-ensemble S de variables appelées *ensemble de coupe* (Loop Cutset) qui en supprimant les arcs qui en sont issus, permet d'obtenir un graphe

réduit qui est un arbre. D'où l'utilisation de l'algorithme des messages locaux. Puis les résultats de toutes les instanciations sont combinés par leurs probabilités à priori. La complexité de cet algorithme est bien exponentielle en fonction de la taille de l'ensemble de coupe.

5.1.4 Algorithme d'Élimination de Bucket

L'algorithme d'élimination de Bucket décrite dans [14] consiste à marginaliser la distribution de probabilité jointe d'un réseau en procédant variable par variable. Chaque marginalisation sur une variable x_i donne lieu à une somme des probabilités de cette variable. Parfois, cette somme vaut 1, ce qui conduira à l'élimination de la variable x_i qui est alors dite inutile. Un ordre des variables doit être donné en entrée et sera l'ordre d'élimination des variables. Trouver le meilleur ordre est un problème NP-difficile. Cette méthode est avantageuse lorsque l'ordre d'élimination des variables est déjà connu où si le réseau est peu dense, mais avec de nombreux circuits. Cet algorithme utilise des potentiels ce qui le rend applicable dans le cas où les nœuds du réseau ne sont pas discrets. L'algorithme se décline ainsi :

1. Choisir un ordre des variables
2. Initialiser la liste des facteurs actifs (les CPDs dans Bayes Net)
3. Présenter les éléments de l'évidence en ajoutant à la liste de facteurs actifs les évidences potentielles.
4. Pour $i = 1, \dots, n$
 - a. Prendre la prochaine variable X_i
 - b. Prendre tous les facteurs qui ont X_i comme argument au large de la liste des facteurs, les multiplier puis additionner sur toutes les valeurs de X_i . un nouveau facteur m_{x_i} est alors obtenu.
 - c. Mettre m_{x_i} dans la liste des facteurs actifs.

Il existe d'autres méthodes d'inférence exacte telles que la méthode d'*inversion d'arcs* introduite par Shachter (1986) [15] qui consiste à changer la direction des arcs en utilisant la loi de Bayes jusqu'à ce que le réseau soit transformé de telle façon que les nœuds de la requête soient des nœuds des enfants observés et la méthode de l'*explication la plus probable (MPE most probable explanation)* dont le principe est de trouver l'état le plus probable. Une technique introduite par Lauritzen & Spiegelhalte se résume à remplacer les signes plus par des max et les signes produit par des min dans les formules de l'inférence classique.

5.2 Méthodes d'inférence approchées

L'inférence exacte dans un réseau bayésien est un problème NP-difficile. Cela peut conduire à des temps de calcul prohibitifs pour des réseaux complexes c'est-à-dire des réseaux ayant un très grand nombre de variables où dont les informations sont incomplètes. En outre, il est impossible de calculer directement la loi de probabilité d'un nœud où d'effectuer une inférence plus complexe. Le seul recours donc est d'introduire un nouveau type d'inférence nommé *inférence approximative* ou *approchée*. Ces méthodes approchées sont principalement de deux types. La première catégorie utilise les méthodes exactes, mais opère seulement sur une

partie du graphe (réseau approché du graphe réel). La seconde catégorie utilise les méthodes stochastiques de simulation. Les méthodes exactes sur des topologies approchées utilisent des réseaux approchés du réseau réel. Ils sont construits en éliminant un certain nombre de connexions du réseau dans le but d'effectuer l'inférence plus rapidement. Elles sont relativement récentes et sont globalement réparties en deux grandes écoles [16]. La première considère l'existence d'une dépendance faible entre deux nœuds du réseau bayésien, cette dépendance est alors dite qualitative. Quantitativement, elle est insignifiante c'est-à-dire que les variables ont un comportement de variables indépendantes. L'idée de l'algorithme de propagation est d'éliminer de tels arcs rendant ces deux nœuds indépendants. De cette façon, les calculs seront accélérés, car la taille des matrices de probabilités conditionnelles en est réduite et l'erreur engendrée reste raisonnable. L'idée de la deuxième école est légèrement différente de celle de la première ; il s'agit de conserver le graphe original, mais de n'effectuer la propagation des informations que partiellement. Plusieurs méthodes ont été proposées : par exemple [17] et [18] réalise la propagation dans la totalité du réseau, mais n'utilise que des sous parties des matrices de probabilités.

En ce qui concerne les méthodes stochastiques, leur principe consiste à accomplir un très grand nombre de tirages aléatoires compatibles à des lois de probabilités conditionnelles et d'estimer, ensuite, la probabilité recherchée définie par des paramètres d'une loi de densité π (moyenne, écart type,...) à l'aide d'une base de données.

6. Apprentissage dans les réseaux bayésiens

L'inférence est un outil qui permet de passer d'un modèle de connaissance et d'une situation à une conclusion. Nous nous intéressons donc dans cette partie à la construction d'un tel modèle. Ce problème est dit problème d'apprentissage, il est séparé en deux parties à savoir l'apprentissage des paramètres qui consiste à estimer des paramètres à partir de données observées et l'apprentissage de la structure qui permet de trouver la meilleure structure par l'apprentissage automatique à partir de données observées.

6.1 Apprentissage des paramètres

Etant donnée une structure de RB G et un ensemble de données D , le principe de l'algorithme d'apprentissage des paramètres est d'estimer les distributions de probabilités (paramètres) $\theta = \{\theta_i\}$ avec $\theta_i = P(X_i = x_k \mid \text{Pa}(X_i) = x_j)$. Cette estimation peut être effectuée différemment selon que les données disponibles soient complètes ou incomplètes.

6.1.1 Apprentissage à partir de données complètes

Apprentissage statistique : C'est la méthode la plus simple et la plus utilisée dans le cas où toutes les variables sont observées. L'estimation statique consiste à estimer la probabilité d'un événement par la fréquence d'apparition de l'événement dans la base de données. Cette approche est appelée maximum de vraisemblance (MV) :

$$P(X_i = x_k \mid \text{Pa}(X_i) = x_j) = \theta_{i,j,k}^{MV} = \frac{N_{i,j,k}}{\sum_k N_{i,j,k}} \quad (24)$$

où $N_{i,j,k}$ est le nombre d'évènements dans la base de données pour lesquels la variable X_i est dans l'état x_k et ses parents sont dans la configuration x_j .

Apprentissage bayésien : Pour cet apprentissage, le principe de l'estimation bayésienne consiste à trouver les paramètres θ les plus probables sachant que les données ont été observées, en utilisant des a priori sur les paramètres. Cette méthode est aussi appelée le maximum a posteriori (MAP) :

$$P(X_i = x_k \mid \text{Pa}(X_i) = x_j) = \theta_{i,j,k}^{MAP} = \frac{N_{i,j,k} + \alpha_{i,j,k} - 1}{\sum_k (N_{i,j,k} + \alpha_{i,j,k} - 1)} \quad (25)$$

où $\alpha_{i,j,k}$ sont les paramètres de la distribution de Dirichlet de la loi a priori $P(\theta)$.

Une autre approche bayésienne consiste à calculer l'espérance a posteriori des paramètres $\theta_{i,j,k}$ au lieu d'en chercher le maximum. Cette approche d'espérance a posteriori (EAP) nous donne :

$$P(X_i = x_k \mid \text{Pa}(X_i) = x_j) = \theta_{i,j,k}^{EAP} = \frac{N_{i,j,k} + \alpha_{i,j,k}}{\sum_k (N_{i,j,k} + \alpha_{i,j,k})} \quad (26)$$

6.1.2 Apprentissage à partir de données incomplètes

Dans la plupart des applications, les données sont incomplètes, *ie.* certaines variables d'intérêt du système à étudier ne sont observées que partiellement ou même jamais, que se soit à cause d'une panne de capteur, d'une variable mesurable seulement dans un contexte bien précis ou d'une personne sondée ayant oublié de répondre à une question, etc.

Le traitement des données manquantes dépend de leur nature. On distingue trois cas : MCAR (Missing Completely At Random) où la probabilité qu'une donnée soit manquante ne dépend pas de l'ensemble de données observées et de données manquantes, MAR (Missing At Random) où la probabilité qu'une donnée soit manquante dépend des données observées et NMAR (Not Missing At Random) où la probabilité qu'une donnée soit manquante dépend à la fois des données observées et manquantes.

Les situations MCAR et MAR sont les plus faciles à résoudre. En effet, pour le premier cas, il n'a besoin d'aucune donnée observée et pour le deuxième cas, il a besoin de données observées qui contiennent toutes les informations nécessaires pour estimer la distribution des données manquantes. La citation NMAR est plus délicate, car il faut d'abord faire appel à des informations extérieures pour réussir à modéliser la distribution des données manquantes et revenir à une situation MCAR et MAR.

La méthode d'estimation la plus utilisée lorsque les variables sont partiellement observées est celle qui se base sur l'algorithme EM (Expectation Maximisation). Nous allons présenter les grandes lignes de cet algorithme dans le cas de l'apprentissage statistique puis de l'apprentissage bayésien.

Apprentissage statistique et algorithme EM

Soient $D_0 = \{X_{0i}^l\}_{i=1..N}$ l'ensemble des données observées et $\theta^{(t)} = \{\theta_{i,j,k}^{(t)}\}$ les paramètres du réseau bayésien à l'itération t .

L'algorithme de maximisation de l'espérance se décline ainsi :

- Calcul de l'espérance : L'espérance est l'estimation des $N_{i,j,k}$ manquantes. Elle est obtenue en calculant leur moyenne conditionnellement aux données et aux paramètres courants du réseau :

$$N_{i,j,k}^* = E [N_{i,j,k}] = \sum_{l=1}^N P(X_i = x_k | Pa(X_i) = C_j, x_0^l, \theta^{(t)}) \quad (27)$$

- Calcul de la maximisation de vraisemblance : La maximisation de vraisemblance est calculée en remplaçant les $N_{i,j,k}$ manquants par leur valeur moyenne calculée précédemment. Les nouveaux paramètres $\theta^{(t+1)}$ obtenus par l'approche de maximum de vraisemblance sont :

$$\theta_{i,j,k}^{t+1} = \frac{N_{i,j,k}^*}{\sum_k N_{i,j,k}^*} \quad (28)$$

- les deux étapes précédentes sont itérées tant que $|\theta^t - \theta^{t-1}| \geq \epsilon$.

Apprentissage bayésien et algorithme EM

L'algorithme EM reste applicable dans le cadre bayésien, en changeant le maximum de vraisemblance de l'étape de maximisation M par un maximum (ou une espérance) à posteriori. Dans ce cas là, l'espérance a posteriori est définie par :

$$\theta_{i,j,k}^{(t+1)} = \frac{N_{i,j,k}^* + \alpha_{i,j,k}}{\sum_k (N_{i,j,k}^* + \alpha_{i,j,k})} \quad (29)$$

$\alpha_{i,j,k} = 1$ pour une distribution de Dirichlet uniforme.

6.2 Apprentissage de la structure

L'apprentissage de la structure d'un RB permet de déterminer automatiquement une structure optimale d'un RB à partir de l'information contenue dans les données observées. Cependant, la recherche dépend fortement du nombre de variables, d'arcs et de valeurs.

On distingue trois grandes familles d'approches d'apprentissage de la structure :

1. Les méthodes basées sur les contraintes et recherche d'indépendances (constraint-based).
2. Les méthodes basées sur le score et la recherche (score and search based)
3. Les méthodes hybrides.

6.2.1 Méthodes basées sur les contraintes et recherche d'indépendances

Cette première série d'approches d'apprentissage de structure, souvent appelée recherche sous contraintes, est issue des travaux de deux équipes « concurrentes », Pearl et Verma d'une part avec les algorithmes IC et IC*, Spirtes, Glymour et Scheines et d'autre part, avec les algorithmes SGS, PC, CI, FCI [19]. Ces algorithmes sont tous basés sur un principe identique dont les étapes sont :

- construire un graphe non dirigé contenant les relations entre les variables, à partir de tests d'indépendance conditionnelle.

- Détecter les V-structures (en utilisant aussi des tests d'indépendance conditionnelle).
- Propager les orientations de certains arcs.
- Prendre éventuellement en compte les causes artificielles dues à des variables latentes.

6.2.2 Méthodes basées sur l'optimisation d'un score

On utilise des algorithmes à base des scores pour trouver la configuration optimale d'un réseau bayésien. Ces algorithmes consistent à parcourir tous les graphes possibles puis associer un score à chaque graphe. Le graphe qui possède le plus grand score va être sélectionné. Cependant, cette méthode est applicable seulement pour les populations à taille limitée (quelques centaines de variables), ce problème dû à la taille super-exponentielle de l'espace de recherche en fonction du nombre de variables.

Chaque RB $B = (G, \theta)$ est noté par un score $S(G, D)$ qui désigne sa capacité à représenter les données D . Ce score doit prendre en compte deux propriétés :

- Décomposable localement, c'est-à-dire qu'il peut s'écrire comme une somme ou un produit de mesures dont chacune n'est fonction seulement que d'un nœud et de l'ensemble de ses parents :

$$S(G, D) = \sum_{X_i \in X} s(X_i, Pa(X_i)) \quad \text{ou} \quad S(G, D) = \prod_{X_i \in X} s(X_i, Pa(X_i))$$

- Equivalent, c'est-à-dire qu'il faut assurer que les réseaux équivalents ont le même score.

La plupart des scores existants dans la littérature appliquent le principe de *parcimonie du rasoir d'Occam*, c'est-à-dire de trouver le modèle qui correspond le mieux aux données D mais qui soit le plus simple possible. Ces scores sont souvent décomposables en deux termes : un premier terme qui est la vraisemblance $L(D|\theta, B)$ et un second terme qui va tenir compte de la complexité du modèle. Les scores les plus utilisés sont : AIC (Akaike's Information Criterion) [20], BIC (Bayesian Information Criterion) [21] et les scores bayésiens BDe (Bayesian Dirichlet equivalent) [22].

6.2.3 Méthodes hybrides

Le but de cette méthode est de combiner les avantages des méthodes basées sur la recherche d'indépendance et celle basée sur les scores. Elles consistent en deux étapes : une recherche locale et une optimisation globale. La recherche locale permet d'obtenir un voisinage contenant toutes les dépendances (indépendances) locales intéressantes avec l'aide de tests d'indépendances. Citons par exemple MMMB (Max Min Markov Blanket) et MMPC (Max Min Parents and Children) [38]. L'optimisation globale permet de faire une recherche sur l'espace des DAG en se restreignant aux dépendances locales trouvées précédemment. Citons par exemple MMHC [23] qui combine MMPC et la recherche gloutonne.

6.2.4 Quelques algorithmes d'apprentissage de la structure

- *Les algorithmes PC et IC (algorithmes à base des tests statistiques)*

La méthode de l'algorithme PC (pour Peter and Clark, les prénoms des inventeurs de la méthode) est une des premières méthodes de recherche

d'indépendance, elle a été introduite par spirtes, Glymour et Scheines en 1993 [19]. L'algorithme PC utilise un test statistique (X^2 ou G^2) pour évaluer s'il y a indépendance conditionnelle entre deux variables ou non. A partir de l'ensemble des tables d'indépendances conditionnelles, il est alors possible de reconstruire la structure du réseau bayésien. En pratique, un graphe complètement connecté sert de point de départ et lorsqu'une indépendance conditionnelle est détectée, l'arc correspondant est retiré. Un algorithme de principe similaire (IC, pour Inductive Causation) a été proposé par pearl [24]. La principale différence entre ces deux algorithmes est que l'algorithme PC est initialisé avec un graphe complètement connecté et raisonne par cartes d'indépendances successives jusqu'à obtenir une I-map[25] minimale tandis que l'algorithme IC est initialisé avec un graphe vide et raisonne par cartes de dépendances successives jusqu'à obtenir une D-map[25] maximale.

- *La recherche gloutonne (algorithme à base des scores)*

L'algorithme de recherche gloutonne (ou GS pour greedy search) est très répandu en optimisation. Il prend un graphe de départ, définit un voisinage de ce graphe, puis associe un score à chaque graphe du voisinage. Le meilleur graphe est alors choisi comme point de départ de l'itération suivante. Le voisinage d'un DAG est défini par l'ensemble de tous les DAG obtenus soit en ajoutant, soit en supprimant, soit en retournant un arc du graphe d'origine. L'algorithme s'arrête lorsque le graphe obtenu réalise un maximum (local) de la fonction de score.

7 Extensions de réseaux bayésiens

7.1 Réseaux bayésiens multi-agents

Un réseau bayésien multi-agent est un réseau partagé par plusieurs agents (ordinateurs, entreprises,...etc). Ils sont utiles lorsque les informations ne sont disponibles que localement. En effet, chaque agent ne connaît qu'une partie de la structure et peut y faire de l'inférence localement. Un exemple de réseau bayésien multi-agent (appelé également Multi-sectioned Bayesian Network) représenté dans la figure 9.

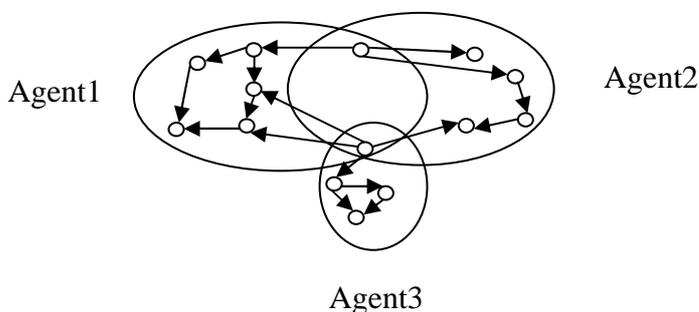


Fig.9 Exemple de réseau bayésien multi agents (RBMA)

7.2 Réseaux bayésiens orientés objets

Un réseau bayésien orienté objets est un réseau bayésien dont une sous-structure (objet) et ses paramètres apparaissent de façon répétée. Il existe plusieurs types d'implémentation pour ces modèles [26] ou [27]. Il permet également de modéliser des systèmes complexes pour lesquels le même mode de raisonnement apparaît dans différents sous-systèmes.

7.3 Diagrammes d'influence

Le diagramme d'influence est un réseau bayésien contenant des nœuds d'utilité et des nœuds de décision dans le but de remettre la modélisation d'un système sur lequel l'opérateur peut agir. En effet le décideur souhaite toujours associer une valeur d'utilité à chaque décision possible. Cette utilité représente la qualité où le coût lié à ces décisions. Ces modèles ont été initialement introduits comme une représentation compacte des graphes de décision par [28][29]. Ils sont à présent plutôt vus comme des extensions de réseaux bayésiens. Les nœuds probabilistes sont représentés par des cercles, les nœuds décision par des carrés et les nœuds utilité par des losanges (fig.10)

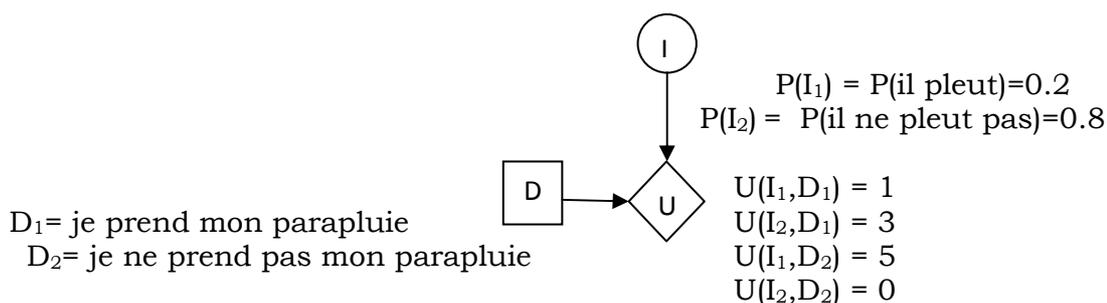


Fig.10 Exemple d'un diagramme d'influence

Dans cet exemple, l'objectif est de minimiser l'espérance d'utilité U en prenant la bonne décision D en fonction de l'état du nœud aléatoire I

7.4 Réseaux bayésiens dynamiques

les réseaux bayésiens dynamiques (DBN) sont une extension des réseaux bayésiens statique dans laquelle l'évolution temporelle des variables $X(t) = (x_1(t), x_2(t), \dots, x_n(t))$ est représentée. Par ailleurs, Un réseau bayésien représente la distribution de probabilité jointe de ces variables dans un intervalle de temps $[0, T]$. Cette JPD est factorisée comme suit :

$$P[X(0), \dots, X(T)] = \prod_{t=0}^T \prod_{i=1}^n (P(x_i(t) / \text{Pa}(x_i(t))) \quad (30)$$

La structure d'un DBN peut être vue comme une concaténation de plusieurs BN_s statiques liés avec les arcs temporels. Nous appelons chacun de ces réseaux comme une tranche de temps (une tranche de temps est définie comme une collection de l'ensemble de variables $X(t)$ dans un seul temps t instantané et leurs parents $\text{Pa}(x(t))$ associés dans la structure graphique). Si le processus est stationnaire, les

hypothèses d'indépendance et les probabilités conditionnelles sont identiques pour tous les temps. Dans ce cas, le DBN peut être représenté par un BN dont la structure est dupliquée pour chaque pas de temps. De ce point de vue, il est évident qu'un *Modèle de Markov caché* (Hidden Markov Model HMM) est un cas particulier de DBN comme le montre la figure 11. Le nœud x_t (respectivement O_t) est une variable aléatoire dont la valeur indique l'état occupé (respectivement l'observation) au temps t .

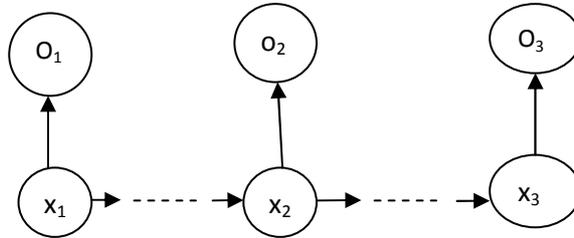


Fig.11 HMM sous forme de réseau bayésien

Les réseaux bayésiens peuvent aussi être représentés sous forme de filtres. Un filtre bayésien peut être représenté par un DBN qui possède qu'une variable d'état et une variable d'observation. Soit x_t la variable d'état et O_t celle d'observation au temps t . Alors pour un filtre bayésien, nous supposons que nous avons la décomposition de la loi jointe à chaque instant t :

$$P(x_1, \dots, x_t, O_t) = P(x_1) \prod_{i=2}^t p(x_i / x_{i-1}) p(O_t / x_t) \quad (31)$$

L'un des filtres les plus utilisés *le filtre de Kalman* qui est un modèle de Markov dont l'à priori initial est gaussien, les variables d'état sont continues et les modèles de transition et d'observation sont gaussiens. Ce modèle est également appelé système dynamique linéaire. Maybeck (1979)[30] et Anderson & Moore (1979)[31].

Il existe des réseaux bayésiens partiellement dynamiques [32] qui comme leur nom l'indique peuvent contenir des nœuds statiques et des nœuds dynamiques.

7.5 Les réseaux bayésiens adaptés à la classification

Dans les réseaux bayésiens, toutes les variables sont considérées de manière identique. Or, pour une tâche de classification, il est pertinent de considérer différemment le nœud représentant la classe. Dans ce qui suit, nous introduisons quelques types de réseaux bayésiens adaptés à la classification.

7.5.1 Les réseaux bayésiens naïfs

Ces réseaux ont une structure simple et unique qui se compose de deux niveaux seulement : Le premier contient un seul nœud Parent et le second contient plusieurs nœuds enfants. La classification est assurée en considérant le nœud parent comme une variable non observée précisant à quelle classe appartient chaque objet et les nœuds enfants comme étant des variables observées correspondant aux différents attributs spécifiant cet objet. Cette structure suppose que toutes les variables d'observation sont indépendantes deux à deux conditionnellement à la classe ce qui entraîne la simplification de la loi jointe ci-

dessous. Supposons que l'on dispose de n variables aléatoires $(X = x_1, \dots, x_n)$ avec C la variable aléatoire représentant la classe (12).

$$P(C, x_1, x_2, \dots, x_n) = P(C) P(x_1/C) \dots P(x_n/C) \quad (32)$$

Ainsi, on obtient la formule suivante :

$$P(C/x_1, x_2, \dots, x_n) = \frac{P(x_1, x_2, \dots, x_n, C)}{P(x_1, x_2, \dots, x_n)} \quad (33)$$

$$= \frac{P(C) P(x_1/C) \dots P(x_n/C)}{P(x_1, x_2, \dots, x_n)}$$

Cette structure pourtant très simple donne de très bon résultats dans de nombreuses applications

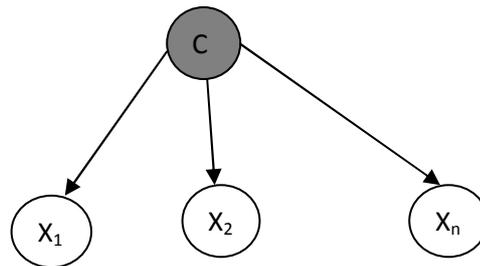


Fig.12 Réseau bayésien naïf

7.5.2 Réseaux bayésiens naïfs augmentés (TANB)

L'hypothèse d'indépendance conditionnelle des variables feuilles est largement non vérifiée. Dans le but d'améliorer les caractéristiques de classification par ce type de réseau bayésien, il a été proposé d'augmenter la structure naïve en ajoutant des arcs entre certaines caractéristiques (feuilles) (fig.13). Une des méthodes proposée pour augmenter le réseau bayésien naïf est TANB (Tree Augmented Naïve Bayes).

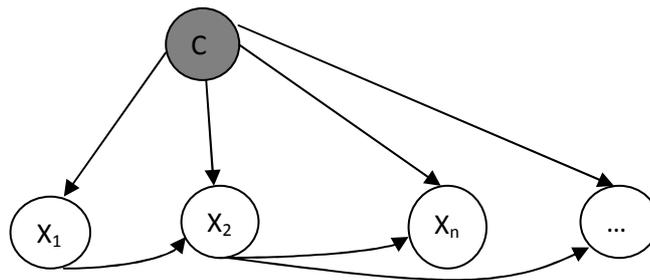


Fig.13 Réseau bayésien naïf augmenté

7.5.3 Les multi-nets bayésiens

C'est une généralisation des structures de Bayes naïves augmentées dans le sens où elles autorisent différentes relations entre les attributs pour les différentes valeurs de la classe. Les classifieurs construits en utilisant ces structures permettent alors d'obtenir de meilleurs résultats que l'utilisation de classifieurs de Bayes naïfs.

7.6 Réseaux bayésiens causaux

Un réseau bayésien est dit causal si tous les arcs représentent des relations de causalité. En effet, une modélisation suppose une connaissance parfaite du problème et de toutes les causes envisageables. Mais il n'est pas toujours possible d'avoir une connaissance parfaite des variables pouvant entrer dans les relations de causalité. Parmi ces réseaux bayésiens, on trouve les modèles causaux markoviens, les modèles causaux semi-markoviens et les graphes ancestraux maximums. Notons qu'un réseau causal markovien (ou réseau bayésien causal) est un graphe orienté sans circuit pour lequel tous les arcs représentent des relations de causalité. De plus, l'ensemble des variables vérifie l'hypothèse de suffisance causale. Ce type de réseau bayésien nécessite la connaissance de toutes les variables pertinentes pour la modélisation. Néanmoins, il est possible que certaines de ces variables soient cachées. Quant aux modèles causaux semi-markoviens, ce sont des graphes causaux sans circuit contenant, à la fois, des arcs dirigés et des arcs bi-dirigés. Les nœuds du graphe représentent les variables observables (x_i) et les arcs bi-dirigés représentent implicitement les variables latentes (l_i). Dans un modèle markovien semi-causal, toute variable latente ne possède pas de parents et possède exactement deux variables observables filles. De plus, il ne vérifie pas l'hypothèse de suffisance causale. En ce qui concerne, le graphe ancestral (sans conditionnement), il s'agit d'un graphe sans circuit contenant des arcs dirigés et des arcs bi-dirigés de telle manière qu'il n'existe pas d'arcs bi-dirigés entre deux variables qui sont connectés par un chemin dirigé. Ce graphe ancestral est dit maximal si pour toute paire de nœuds (x, y) non adjacents, il existe un ensemble de séparation z tel que $x \perp\!\!\!\perp y / z$ [33].

Notons que tout modèle causal avec des variables latentes peut être transformé en modèle causal semi-markovien. Cette transformation sauvegarde les relations d'indépendance et de causalité entre les variables observées (fig.14).

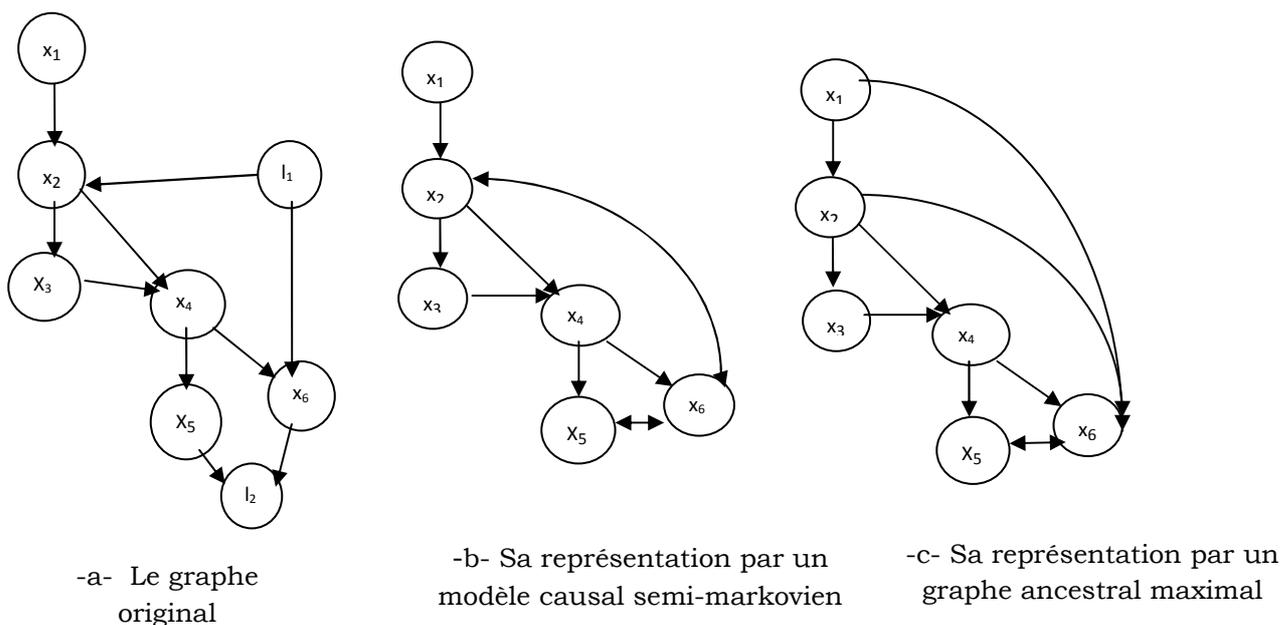


Fig.14 différents types de modélisation par variables latentes

D'autres variantes de réseaux bayésiens existent tel que *les réseaux bayésiens à temps continu* qui permettent la modélisation des processus stochastiques structurés qui évoluent continuellement dans le temps [34], *les réseaux bayésiens multi-entités* qui unifient la logique du premier ordre et la théorie des probabilités [35]. Une autre version de réseaux bayésienne associe la théorie des possibilités au lieu de celle des probabilités à la logique floue appelée *les réseaux possibilistes* [36]. Il est toujours possible de mélanger différentes représentations ou interprétations des réseaux bayésiens. Par exemple, Maes, Meganck et Manderick [37] et Maes [38] introduisent des réseaux bayésiens multi-agents causaux.

8 Réseaux bayésiens à densité de probabilité continue

Les réseaux bayésiens, tels que présentés dans ce travail, comportent des variables prenant leurs valeurs dans des espaces discrets. Bien qu'il soit usuel de travailler sur de tels modèles, essentiellement pour des raisons pratiques, il est tout à fait possible d'employer des réseaux bayésiens dans le cas où les variables modélisées sont continues. Ainsi, [39] ont proposé des réseaux bayésiens dont les variables présentent une densité de probabilités correspondant à une distribution gaussienne. D'autres modélisations permettent de généraliser la densité modélisée en l'approximant par un mélange de gaussiennes [40] ou encore des densités exponentielles tronquées [41]. Les modèles utilisant les distributions gaussiennes sont les plus utilisés. La distribution d'un nœud conditionnellement à ses parents est une distribution gaussienne est une combinaison linéaire dont la moyenne est la valeur des parents, et dont la variance est indépendante de la valeur des parents. La distribution d'un nœud conditionné à ses parents est donc :

$$p\left(X_i = x \mid \Pi_i = \pi\right) = N(m_i, v_i) \quad (34)$$

$$m_i = \mu_i + \sum_{j=1}^{n_i} \beta_{ij} (x_j - \mu_j) \quad (35)$$

$$\beta_{ji} = \frac{\sigma_{ij}}{\sigma_j^2} \quad \text{et} \quad v_i = \sigma_i^2 \quad (36)$$

où :

Π_i Sont les parents de X_i

μ_i et μ_j sont les espérances des attributs X_i et X_j respectivement sans prendre en compte les parents.

m_i et v_i : sont la moyenne et la variance de l'attribut X_i en connaissant ses parents X_j .

n_i : est le nombre de parents de X_i .

β_{ij} : Appelé matrice de régression des poids ou matrice de poids.

9. Conclusion

Nous avons abordé dans ce chapitre les fondements théoriques des réseaux bayésiens. Nous avons également présenté un bref état de l'art des mécanismes d'inférence et d'apprentissage nécessaires à la mise en place d'un réseau bayésien ainsi que des extensions de ce dernier. En effet, le réseau bayésien est un graphe acyclique orienté dont les paramètres sont les probabilités conditionnelles et la structure du réseau. Lors de la construction de réseaux bayésiens, il n'est pas toujours évident qu'un expert puisse fournir de façon numérique l'ensemble des paramètres nécessaires à l'inférence dans un graphe. Il est intéressant dans certains cas de déterminer ces paramètres à partir d'une base d'exemples. C'est grâce à la méthode d'apprentissage qui permet de remédier à ce problème. Les deux méthodes utilisées sont *apprentissage de paramètres* et *apprentissage de structure*.

Une fois qu'un réseau bayésien construit pour rendre compte d'un domaine, on cherche souvent à l'utiliser pour déterminer des probabilités correspondant à certains événements, certaines questions, certaines dépendances. Dans ces cas, les inférences permettent effectivement de calculer la probabilité de chaque nœud lorsque les autres variables sont inconnues. Il existe cependant quelques méthodes permettant d'effectuer ce calcul qui se divisent en deux catégories : méthodes d'inférence exactes et les méthodes d'inférence approchées.

Ce chapitre nous permet maintenant à mieux distinguer les avantages et les inconvénients que peut avoir un modèle graphique probabiliste face à un modèle graphique standard. En effet, les réseaux bayésiens possèdent tous les avantages des modèles probabilistes standards, ainsi que des avantages supplémentaires liés à leur représentation graphique. Ils offrent un outil naturel pour résoudre deux problèmes qui apparaissent tout le temps dans les domaines de mathématiques appliquées et de l'ingénierie - l'incertitude et la complexité - et jouent un rôle de plus en plus important dans la conception et l'analyse d'algorithmes d'apprentissage par machine (machine learning algorithms).

Le prochain chapitre sera consacré à appliquer les réseaux bayésiens dans le cadre du traitement d'image pour la détection de la peau.

Chapitre 3

Détection de la peau

1 Introduction

La détection de la peau consiste à différencier les pixels de peau de ceux de non peau. Elle constitue une phase primordiale de prétraitement dans plusieurs applications telles que le suivi des gestes de la main [42] [43], la vidéo surveillance [44] [45], le filtrage internet [46][47] et la segmentation de la vidéo[48][49]. 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 peau de ceux de non-peau. L'image en sortie du détecteur possède les mêmes dimensions que celle de l'image originale. La détection de la peau est toute fois une tâche efficace et peu coûteuse, mais difficile à accomplir étant donné la diversité ethnique avec des teintes qui varient selon les personnes et les différentes races (Fig(1)), la variété des conditions de prise de vue telles que les diverses conditions d'éclairage (distribution, orientation et intensité de la source) avec différents dispositifs ayant des caractéristiques spécifiques (capteur, optique, bruit, contraste,...) sans oublier qu'un décor complexe peut compliquer cette tâche (Fig(2)) et Fig(3))



Fig (1) Diversité ethniques



Fig (2) Différents éclairages

Fig (3) Décors complexes

2 Les méthodes de détection de la peau

Il existe de nombreuses méthodes de détection des régions de peau dans une image couleur. On en identifie trois approches majeures: l'approche basée sur la géométrie et les traits caractéristiques généralement utilisées pour segmentation des régions de peau du visage et/ou des mains dans le but de reconnaissance de gestes d'individu, approche basée sur le mouvement utilisée pour segmentation de régions de peau dans des scènes vidéo, et l'approche basée sur la couleur que nous développons dans ce qui suit.

2.1 Approche basée sur la couleur [49][50]

L'exploitation de l'information de couleur est également un moyen efficace pour la discrimination entre les pixels de peau et ceux de non-peau. Ce qui rend cette approche facile à exploiter est le fait que la couleur de la peau est regroupée dans une bande étroite de l'espace de couleur. Dans une approche basée sur la couleur, deux axes de recherche sont abordés : le choix de l'espace couleur que nous devons adopter pour une bonne classification et la méthode de modélisation de la couleur permettant de distinguer les pixels de peau de ceux de non-peau. La couleur est la primitive la plus répandue pour la séparation des pixels peau des pixels non-peau dans une image. La raison principale du choix de ce descripteur

revient au fait que la gamme des couleurs de la peau des êtres humains est située dans une bande étroite de l'espace de couleur grâce à sa constitution d'une combinaison de sang (rouge) et de mélanine (brun, jaune). Des études ont montré que la variabilité de la couleur de la peau tenait plus de la différence d'intensité plutôt que de la chrominance. Ainsi, les chercheurs ont essayé d'utiliser des espaces dans lesquels la couleur peau resterait idéalement invariante aux conditions d'éclairage. Cependant, la couleur de la peau peut être présentée dans différents espaces couleurs tels que l'espace RGB, YCbCr et HSV. Dans l'espace RGB, le changement de luminance d'un patch de peau donné affecte les trois composantes R, G et B. Autrement dit, l'emplacement d'un patch de peau donné dans le cube RGB change en fonction de l'intensité de l'éclairage (luminosité) sous lequel les images de ce patch ont été prises. En dépit de ces limites, l'espace RGB est largement utilisé dans la littérature pour la détection de la peau grâce à sa simplicité. Dans l'espace YCbCr, l'emplacement de la couleur peau dans les deux canaux de chrominance (Cb et Cr) ne sera pas affecté par le changement de l'intensité (luminosité) de l'éclairage de la scène, seuls ces deux canaux sont retenus à l'application de détection de la peau. La séparation explicite des composantes luminance et chrominance rendent cet espace couleur attirant pour la modélisation de la couleur de la peau [51], [42], [53]. L'espace HSV permet à l'utilisateur de spécifier la classe couleur peau en fonction de la teinte et la saturation. Il a été utilisé dans plusieurs travaux de recherches faisant appel à la détection de la peau grâce à la notion de chrominance et de luminance. Comme dans l'espace YCbCr, seuls les deux canaux H et S sont retenus à l'application de détection de la peau.

La détection de la peau utilisant l'approche basée sur la couleur est affectée par la variation des conditions d'éclairage et les objets ayant une couleur semblable à celle de la peau tels que le cuir, le bois, le sable, certains vêtements, etc. Un champ fertile de méthodes utilisées dans la détection de la peau est exposé dans la littérature. Des études et des synthèses sur les techniques de modélisation sont présentées [54][55][56][57][58]. Ces différentes méthodes ont pour objectif de déterminer la distribution de la classe peau et de la classe non-peau. Elles sont réparties en trois divisions : l'ensemble des méthodes basées sur des seuils fixés de définition de la peau, l'ensemble des méthodes paramétriques et l'ensemble des méthodes non paramétriques.

2.1.1 Les méthodes basées sur des seuils fixés de définition de la peau

Une méthode de construction d'un classifieur de peau est de définir explicitement (à partir d'un nombre de règles) les frontières qui délimitent le

faisceau de la peau dans certains espaces couleurs. Par exemple, d'après [63], un pixel d'une image est un pixel peau si :

- Dans l'espace RGB :

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

- Dans l'espace YCbCr :

$$77 \leq Cb \leq 127 \text{ et } 133 \leq Cr \leq 173$$

- Dans l'espace HSV :

$$0 \leq H \leq 50 \quad \text{et} \quad 0.23 \leq S \leq 0.68$$

La simplicité de cette méthode a attiré beaucoup de chercheurs [63], [64], [65] et [66][67]. Son principal avantage est la simplicité des règles de détection de la peau qui mènent à la construction d'un classifieur très rapide. L'importante difficulté en réalisant des taux de reconnaissance hauts en utilisant cette méthode réside dans le choix de l'espace couleur ainsi que les règles de décisions par des études empiriques.

2.1.2 Les méthodes paramétriques

Les modèles de peau paramétriques permettent d'ajuster les distributions avec quelques fonctions spécifiques paramétrées. Différentes fonctions peuvent être appliquées en fonction du problème. Nous présentons dans la suite les méthodes paramétriques les plus utilisées pour construire un modèle de peau et de non-peau.

Modèle basé sur une simple gaussienne

Le modèle paramétrique gaussien est basé sur l'utilisation d'un modèle gaussien définissant la distribution de la couleur de peau dans un espace couleur donné. Par ailleurs, toute distribution d'une classe peut être modélisée par une fonction gaussienne ou par un mélange de gaussiennes. En particulier, dans le cas d'une simple gaussienne, la distribution conditionnelle peau s'écrit :

$$P(c/peau) = \frac{1}{2\pi|\Sigma_{peau}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(c - \mu_{peau})^T \Sigma_{peau}^{-1}(c - \mu_{peau})\right] \quad (1)$$

$$\Sigma_{peau} = \frac{1}{N_{peau}-1} \sum_c N_{peau}(c) (c - \mu_{peau})(c - \mu_{peau})^T \quad (2)$$

où c est la variable aléatoire à deux dimensions représentant le vecteur couleur. L'espérance μ_{peau} et la matrice de covariance Σ_{peau} représentent les paramètres du modèle gaussien. Ces paramètres peuvent être estimés à

partir des exemples constituant la base d'apprentissage. Cette méthode permet de modéliser la distribution de la peau et celle de non-peau. Une fois les distributions conditionnelles de peau et de non-peau sont déterminées, la formule de Bayes appliquée à ces deux distributions permet de calculer la probabilité $P(\text{peau}/c)$.

Modèle basé sur un mélange de gaussiennes

Ce modèle est la généralisation du gaussien simple qui traduit le caractère multimodal de la distribution de la couleur de la peau. Dans ce cas, la densité de probabilité gaussienne est représentée comme suit :

$$P(c/\text{peau}) = \sum_{i=1}^k w_i p_i(c/\text{peau}) \quad (3)$$

où :

-K :le nombre de composants du mélange

- w_i : les coefficients

- p_i : le noyau d'une gaussienne défini par l'équation (3)

Les paramètres à estimer pour la distribution de peau modélisée par un mélange de gaussienne sont $(\mu_{\text{peau}}, \Sigma_{\text{peau}}, w_i)$. Généralement l'algorithme EM (expectation Maximisation) est appliqué pour estimer ces trois paramètres.

Modèle elliptique de borne

D'après Lie et Yoo [59] la distribution de la couleur de peau et de non-peau dans n'importe quel espace couleur prend approximativement la forme d'une ellipse. Ils ont séparé les régions de peau et de non-peau par une borne elliptique afin de caractériser la vraie forme de la distribution des données d'apprentissage. Le modèle est défini comme suit :

$$\Phi(c) = (c - \Phi)^T \Sigma_{\text{peau}}^{-1} (c - \Phi) \quad (4)$$

où Φ et Σ_{peau} sont calculés à partir des données d'apprentissage comme

$$\text{suit :} \quad \Phi = \frac{1}{|c_{\text{peau}}|} \sum_c c_{\text{peau}} c \quad (5)$$

$$\Sigma_{\text{peau}} = \frac{1}{N_{\text{peau}}} \sum_c c_{\text{peau}} N_{\text{peau}}(c) (c - \mu_{\text{peau}}) (c - \mu_{\text{peau}})^T \quad (6)$$

où c_{peau} est le nombre de pixels de couleur de peau avec $C_{\text{peau}} \leq C$. μ_{peau} est l'espérance des vecteurs d'apprentissage de pixels peau défini par l'équation (ci-dessus).

Pour la classification des pixels, $\Phi(c)$ doit être comparé à un seuil fixe (θ) ; c est considéré comme un pixel peau si $\Phi(c) < \theta$ sinon c'est un pixel non peau.

2.1.3 Méthodes non paramétriques

L'idée principale de la modélisation non paramétrique de la peau est d'estimer la distribution de couleur de la peau des données d'apprentissage sans aucune hypothèse sur cette distribution.

Table de correspondance (lookup table LUT)

Plusieurs algorithmes de détection de la peau utilisent des histogrammes pour segmenter les pixels de la peau. Cette méthode consiste à construire dans un premier temps un modèle de peau en se basant sur le calcul d'histogramme à partir d'exemples. Ces histogrammes représentent les couleurs de l'image de traitement et ils sont construits à partir d'un ensemble d'images d'apprentissage contenant les échantillons de la peau sous ses différentes couleurs (blanche, brune, noire, rouge, etc.). Par ailleurs, l'espace couleur est quantifié en un nombre de cases (bins). Chaque case de l'histogramme encode la valeur d'une couleur et son nombre de répétitions dans la base d'apprentissage. Ces cases forment donc un histogramme de couleurs désigné dans certains travaux sous le nom de *la table de correspondance* « *Lookup table* ». La dimension de l'histogramme dépend en partie de la dimension de l'espace de couleur utilisé. On distingue deux types de tables normales de correspondance : 3D et 2D. Nous signalons que les histogrammes (2D) sont les plus souvent utilisés ; nous citons par exemple l'espace (r,g) (composantes R et G normalisées) par Becard [60] utilisant la teinte et la saturation. Quant aux auteurs [61][62], ils ont optés pour l'utilisation de l'espace couleur RGB donc pour des histogrammes 3D. Après construction de l'histogramme, il est normalisé et ses valeurs sont converties en distribution discrète $P(c/peau)$:

$$P(c/peau) = \frac{Peau(c)}{Norm} \quad (7)$$

où $peau(c)$ est le nombre de pixels associés à une case (bin) de l'histogramme de peau formé par le vecteur de couleur c et $Norm$ le coefficient de normalisation qui peut être la somme de toutes les valeurs des cases de l'histogramme ou la valeur maximale des cases.

Les valeurs normalisées des cases de la table de correspondance constituent la vraisemblance que les couleurs de correspondance correspondront à la peau.

Carte auto organisatrice (SOM)

Une SOM dite encore réseau de *Kohonen* est un algorithme d'apprentissage compétitif non supervisé à partir d'un réseau de neurones artificiels (Kohonen, 1984, 2001). Lorsqu'une observation est reconnue, l'activation d'un neurone du réseau, sélectionné par une compétition entre les neurones, a pour effet le renforcement de ce neurone et l'inhibition des autres (c'est la règle du "Winner Takes All"), ce neurone est alors dit neurone *gagnant*. Chaque neurone se spécialise donc au cours de l'apprentissage dans la reconnaissance d'un certain type d'observations. Ces neurones sont donc modifiés en fonction des observations. La structure du réseau de Kohonen est constituée de deux couches. La première,

appelée couche d'entrée, est composée d'un certain nombre de neurones représentant les attributs d'une observation. La couche de sortie, ou couche compétitive est également composée d'un certain nombre de neurones régulièrement répartis sur une carte. Chaque neurone de la première couche est connecté à un neurone de la deuxième couche et chaque connexion possède un vecteur poids. Chaque neurone de la couche de sortie est donc caractérisé par sa position relative sur la carte et par son vecteur poids dans l'espace de représentation des observations. Chaque neurone est associé à ses voisins et la connexion entre ces neurones voisins peut se faire d'une manière rectangulaire ou hexagonale (figure (3.5)). Dans [73], l'algorithme SOM est utilisé pour la détection de la peau. Les auteurs ont construit deux SOMs dont l'un modélise la distribution des pixels peau et l'autre des pixels non-peau.

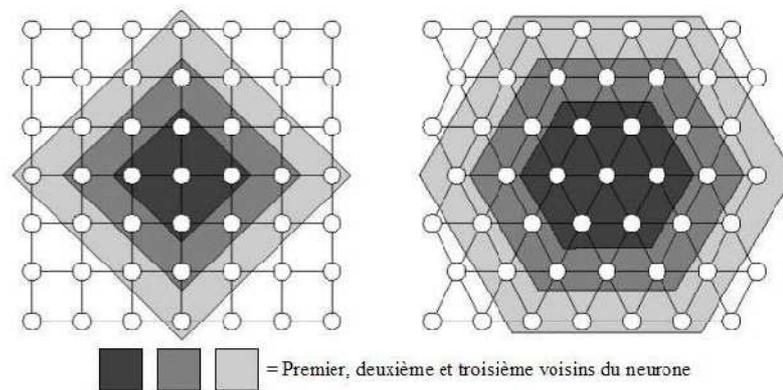


Fig. 4 Disposition des neurones de Kohonen

Il existe d'autres méthodes non paramétriques de la détection de la peau basées sur la décision de Bayes que nous développons dans la section suivante.

3. Classification par la règle de Bayes

3.1 Classifieur bayésien couplé à un histogramme

Dans cette approche, deux histogrammes sont utilisés pour la couleur de la peau et de non peau. Ils quantifient l'espace de couleur C à un certain nombre de cases (bins) $c \in C$ et comptent le nombre de pixels de couleurs dans chaque case. $N_{peau}(c)$ représente donc le nombre de pixels de couleur pour la classe peau et $N_{-peau}(c)$ pour la classe non peau. En fin, ils normalisent chaque bin pour obtenir la distribution conditionnelle des pixels peau $P(c/peau)$ et des pixels non-peau $P(c/non\ peau)$.

Supposons que N_{peau} représente le nombre total de pixels de peau et N_{-peau} le nombre total de pixels non-peau dans la base d'apprentissage. Alors les probabilités conditionnelles peau et non-peau s'écrivent comme suit :

$$p(c/peau) = \frac{N_{peau}(c)}{N_{peau}} \quad (8)$$

$$P(c/-peau) = \frac{N_{-peau}(c)}{N_{-peau}} \quad (9)$$

Ce qui signifie aussi :

$$P(peau) = \frac{N_{peau}}{N_{peau} + N_{-peau}} \quad (10)$$

$$P(-peau) = \frac{N_{-peau}}{N_{peau} + N_{-peau}} \quad (11)$$

$$= 1 - P(peau)$$

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

$$P(peau/c) = \frac{p(c/peau) p(peau)}{p(c/peau) p(peau) + p(c/-peau) p(-peau)} \quad (12)$$

$$P(-peau/c) = \frac{p(c/-peau) p(-peau)}{p(c/peau) p(peau) + p(c/-peau) p(-peau)} \quad (13)$$

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

$$\frac{P(peau/c)}{P(-peau/c)} > \theta, \quad \theta = k \frac{1 - p(peau)}{p(peau)}$$

$P(c/peau)$ et $P(c/-peau)$ sont directement calculés à partir des histogrammes de couleur peau et non-peau. La valeur du seuil θ est calculée empiriquement, et cela, à partir de la courbe ROC (Receiver Operating Characteristic).

Cette méthode est simple et rapide du fait qu'elle ne nécessite que deux histogrammes. Elle est utilisée par plusieurs auteurs [74][75][76][72].

3.2 Classification par réseau bayésien

Pour la modélisation et la classification des pixels peau et non peau, les réseaux bayésiens adaptés à la classification peuvent être utilisés. Parmi ces réseaux les plus populaires le Naïf Bayes (NB) et le Tree Augmented Naïf (TAN) déjà définis dans le second chapitre.

Le NB et le TAN adapté à la classification de pixels peau sont conçus comme des classifieurs à deux classes. En effet, les pixels de l'image seront divisés en deux classes, à savoir une classe « pixels peau » et une classe « pixels non-peau ». Le nœud parent dans les deux cas (BN ou TAN) représente la classe à laquelle appartient chaque pixel de l'image couleur, il est considéré comme une variable non observée. Alors que les nœuds enfants sont des variables observées, correspondant aux intensités des pixels dans l'espace couleur choisi. Les réseaux bayésiens naïfs sont

performants malgré leur simplicité, ils dépassent des techniques beaucoup plus sophistiquées même lorsque l'hypothèse d'indépendance n'est pas toujours vraie. La probabilité a posteriori de chaque classe s'écrit de la manière suivante :

$$P(C_i/X=x_j) = P(C_i) \prod_{j=1}^p P(x_j / C_i) \quad (14)$$

avec :

C_i les valeurs de la classe C telle que, C_1 c'est la classe « pixels peau » et C_2 classe « pixels non peau ».

X_j représente les valeurs de vecteur X de paramètres du réseau qui sont les pixels de l'image

i la i ème classe ; $i=1, 2$

p c'est le nombre de paramètres.

Une fois les probabilités conditionnelles calculées, on affecte chaque pixel à sa classe en appliquant la décision de Bayes suivante :

$$\begin{aligned} d(X) &= \operatorname{argmax}_{\text{classe}} P(C/X) \\ &= \operatorname{argmax}_{\text{classe}} P(X_j / C) P(C) \\ &= \operatorname{argmax}_{\text{classe}} P(C_i) \prod_{j=1}^p P(X_j / C_i) \end{aligned} \quad (15)$$

L'utilisation du modèle Tree-Augmented Naïve Bayes (TAN) a été proposée comme un classificateur, afin d'améliorer la performance sur la simplicité du réseau bayésien naïf malgré que ce dernier donne de meilleurs résultats. Dans la structure du TAN, la variable classe est le parent de tous les nœuds du graphe, et chaque élément a au plus une autre caractéristique de parent, de sorte que le graphe forme un arbre. La structure du TAN est obtenue en maximisant la fonction de vraisemblance compte tenu des données de formation de toutes les structures possibles.

4. Test et résultats

Dans cette partie, nous présentons les résultats de détection de la peau dans une image couleur par le réseau bayésien naïf dans trois espaces couleurs différents (RGB, YCbCr et HSV). Afin de conclure sur le meilleur espace couleur pour la détection de la peau, nous calculons la matrice de confusion qui permet l'obtention du taux de bonne détection pour chaque espace couleur considéré. Dans le but de montrer l'efficacité du classifieur bayésien naïf, nous appliquons la méthode du seuillage aux mêmes images test déjà utilisées du fait qu'elle soit performante pour la détection de la peau. La comparaison entre ces deux méthodes se fait en comparant les taux de bonnes détections, les vrais positifs (VP) et les faux positifs (FP) dans les deux cas considérés tels que, les vrais positifs sont les pixels peau détectés comme des pixels peau, les faux positifs sont les pixels non peau détectés comme des pixels peau, les vrais négatifs sont les pixels

non peau détectés comme des pixels non peau et les faux négatifs sont les pixels peau détectés comme des pixels non peau. Les performances sont exprimées par :

$$\text{Taux de bonne détection} = \frac{\text{nombre de bonnes détections}}{\text{nombre de détections}} \quad (16)$$

$$\text{VP} = \frac{\text{nombre de pixels peau détectés comme peau}}{\text{nombre de pixels peau}} \quad (17)$$

$$\text{FP} = \frac{\text{nombre de pixels non peau détectés comme peau}}{\text{nombre de pixels peau}} \quad (18)$$

La première phase, concernant l'apprentissage, est effectuée en trois étapes :

1. Choisir un espace couleur approprié.
2. Collecter une base de données d'images de peau à partir d'exemples d'images. Une telle base de données contient généralement des patches de peau à partir d'une grande variété de personnes et dans différentes conditions d'éclairage.
3. Apprendre les paramètres du classifieur de peau (le réseau bayésien naïf dans notre cas).

Après l'apprentissage, la phase de classification des pixels de la peau dans une image donnée implique :

1. De convertir l'image dans le même espace couleur que celui qui a été utilisé dans la phase d'apprentissage.
2. De classifier chaque pixel en pixel-peau ou pixel-non-peau, en utilisant le classifieur entraîné.
3. Dans certains cas, on peut appliquer des post-traitements sur les régions détectées comme de la peau en utilisant la morphologie mathématique, et ceci en imposant des critères géométriques et spatiaux à ces régions.

Dans notre application les espaces de représentations suivants : RVB, YCbCr et HSV. Rappelons que la couleur est la primitive la plus répondeuse pour la séparation des pixels peau des pixels non peau dans une image. La raison principale du choix de ce descripteur revient au fait que la gamme des couleurs de la peau des êtres humains est située dans une bande étroite grâce à sa constitution d'une combinaison de sang (rouge) et de mélanine (brun, jaune). Des études ont montré que la variabilité de la couleur de la peau tenait plus de la différence d'intensité plutôt que de la chrominance. Ainsi plusieurs chercheurs ont essayé d'utiliser des espaces dans lesquels la couleur peau resterait idéalement invariante aux conditions d'éclairage.

Nous avons programmé le RBN qu'on utilise pour la détection de la peau sous **MATLAB R2009a**, en utilisant la Bayes Net toolbox de MATLAB (**BNT 2007**). Les informations (régions de peau et régions de non-peau) sur les images tests sont obtenues manuellement en utilisant l'éditeur d'image JIMP.

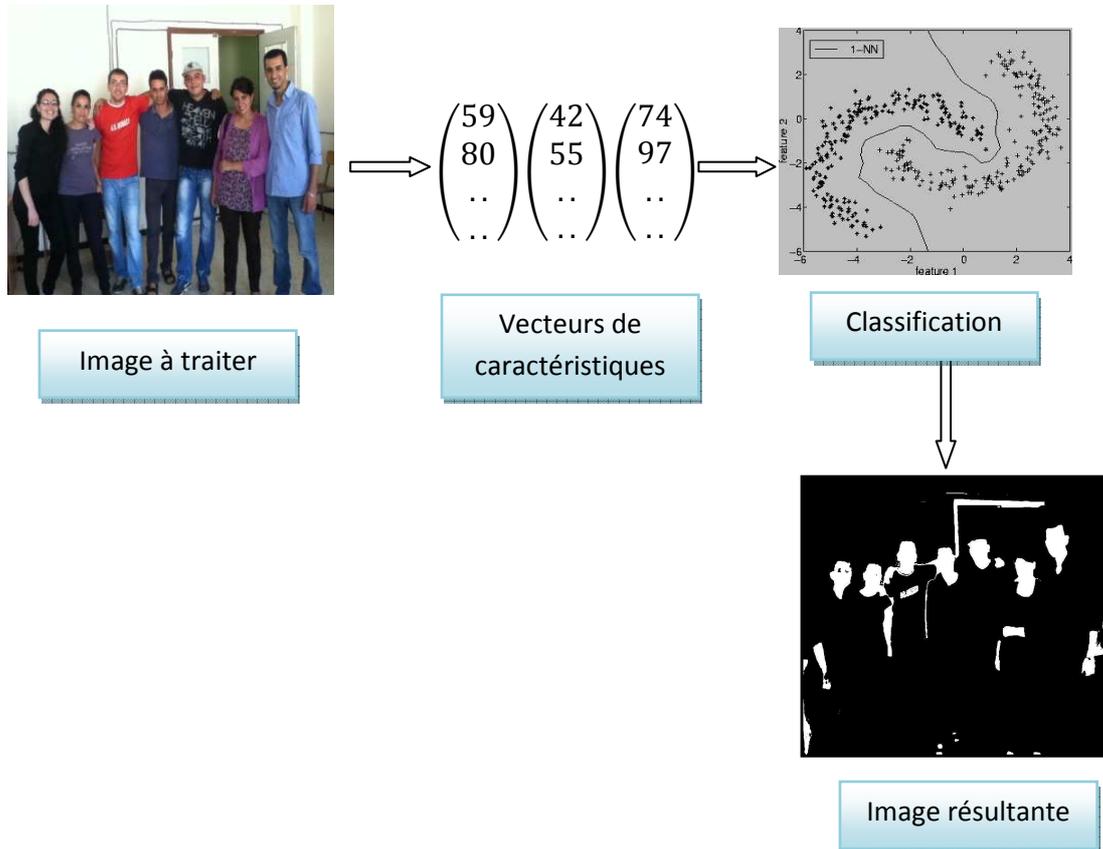


Fig.5 Principe de détection de peau

4.1 Méthode adoptée

Construction de la base de données

Nous avons construit notre propre base de données à l'aide des photos prises pour plusieurs personnes dans différentes conditions d'éclairage. Cette base de données est constituée de 200 images d'apprentissage contenant 100 images peau et 100 images non-peau et quelques images tests contenant des pixels peau et des pixels non-peau. Nous avons ensuite extrait des patches (imassettes) de taille 50×50 à partir des images d'apprentissage dans le but de construire le vecteur des paramètres qui contient les pixels de ces images patch. Toutes les images sont en format JPEG, elles sont nommées par 001.JPG, 002.JPG, etc. La figure (6) représente quelques exemples d'images d'apprentissage et les patches extraits :



Fig.6 Exemple des images d'apprentissage

Création du RBN

Nous créons le réseau bayésien naïf grâce à la commande **mk_bnet**, qui prend en entrée une matrice d'adjacence, la taille des nœuds et les indices des nœuds discrets. Nous allons à présent pouvoir faire de l'apprentissage et de l'inférence dans le réseau crée.

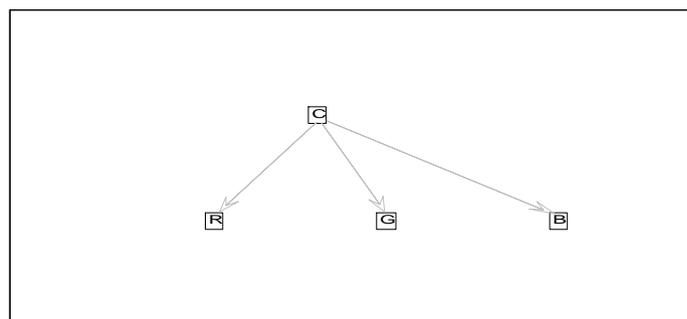


Fig.7 Réseau bayésien naïf obtenu avec les paramètres d'intensités dans le rouge, le vert et le bleu

Les trois nœuds enfants correspondent aux trois composantes de l'espace couleurs RGB qui caractérise chaque pixel de l'image (le rouge R, le vert G et le bleu B). Ces paramètres sont indépendants par rapport au nœud parent qui correspond à la classe C qui représente les pixels peau ou les

pixels non-peau. Dans le cas des images YCbCr et HSV, les nœuds enfants correspondent aux deux composantes chromatiques Cb,Cr pour une image YCbCr et H,S pour une image HSV.

Apprentissage des paramètres

L'apprentissage des paramètres du RBN dans le cas des images RGB et YCbCr dont les variables sont discrètes, est réalisé sans données manquantes par maximum de vraisemblance avec un a priori uniforme de Dirichlet. Cette méthode est implémentée dans BNT sous la fonction **learn_params** qui prend en entrée le réseau créé et la base de données d'apprentissage.

L'algorithme est divisé en deux étapes essentielles :

- Initialiser les probabilités conditionnelles aléatoirement avec un a priori de Dirichlet.
- Calculer les probabilités conditionnelles de chaque attribut par rapport à son parent en utilisant les a priori :

$$P(X=x_j / C=c_i) = \frac{N_{X=x_j, C=c_i} + \alpha_{i,j}}{\sum(N(c_i) + \alpha_{i,j})} \quad (19)$$

où $X = [R,G,B]$, $j = 1, \dots, ns$; ns le nombre de valeurs que peut prendre chaque attribut, $i = 1, 2$ et $C = [c_1, c_2]$ tel que c_1 c'est la classe « pixel peau » et c_2 la classe « pixel non-peau ».

Dans le cas d'une image HSV, les variables des nœuds enfants qui représentent les valeurs de H et de S sont continues, donc nous avons fait appel à la distribution gaussienne pour faire de l'apprentissage des paramètres « **gaussian_CPD** ». Après avoir construit le RBN et calculer toutes les CPDs (les probabilités conditionnelles), nous utilisons ce réseau pour l'inférence.

L'inférence Bayésienne

Une fois que le réseau bayésien naïf est quantifié, nous calculons les probabilités des nouveaux attributs par rapport à la classe peau et à la classe non peau, en utilisant l'équation suivante :

$$P(C/X=x_j) = P(C) \prod_{j=1}^n P(X = x_j / C) \quad (20)$$

Nous avons appliqués l'algorithme de propagation de messages locaux qui est implémenté dans la BNT sous fonction **pearl_inf_engine** dans le cas des images RGB et YCrCb, et l'algorithme d'élimination de variables « **var_elim_inf_engine** » pour l'image HSV. Dans le but de classification de pixels peau et ceux de non peau, nous prenons la probabilité $P(C/X=x_j)$ la plus grande pour définir la classe qui correspond à chaque pixel. Dans la partie suivante, nous exposons les images résultantes en appliquant un RBN aux images tests de la figure (8). La méthode basée sur les seuils fixés de définition de la peau sera effectuée dans les mêmes conditions que la méthode précédente (mêmes images tests et mêmes espaces couleurs).

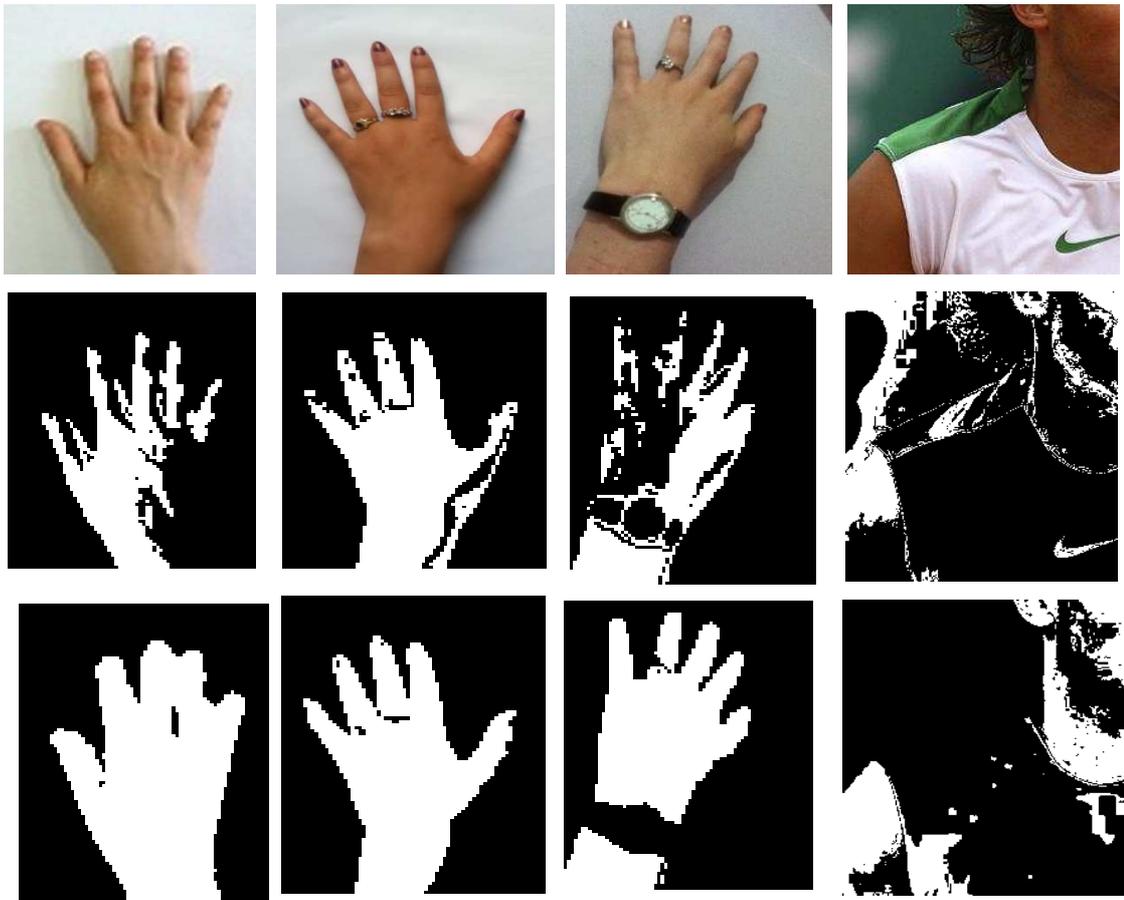
Résultats obtenus par l'espace RGB

Fig. 8 Détection de la peau à partir d'une image RGB, la première ligne représente les images test originales, la deuxième les images résultantes en utilisant un RBN et la troisième les images résultantes en appliquant la méthode des seuils.

Les matrices de confusion en utilisant les deux méthodes considérées, les taux de bonnes détections, les taux de vrais positifs et les taux de faux positifs sont représentés dans le tableau ci-dessous.

Tab.3.1 Résultats obtenus en appliquant le RBN et les seuils pour la détection de la peau à partir des images RGB

	Matrice de confusion	Taux de bonnes détections	VP	FP
Réseau bayésien	$\begin{bmatrix} 4516 & 227 \\ 1062 & 1106 \end{bmatrix}$	81.35 %	51.22 %	10.51%
Seuillage	$\begin{bmatrix} 4554 & 197 \\ 996 & 1164 \end{bmatrix}$	82.73%	53.91 %	9.12%

Le tableau (3.1) montre que l'utilisation de RBN pour la détection de peau à partir des images RGB n'est pas la meilleure méthode par rapport au seuillage, mais la différence est minime (81.35% pour les réseaux bayésiens et 82.73% pour le seuillage de taux de bonne détection). les valeurs des taux de vrais positifs et de faux positifs ne sont pas vraiment importantes (VP de 51.22% et 53.91%, FP de 10.51% et 9.12% pour la méthode des réseaux bayésiens et de seuillage respectivement).

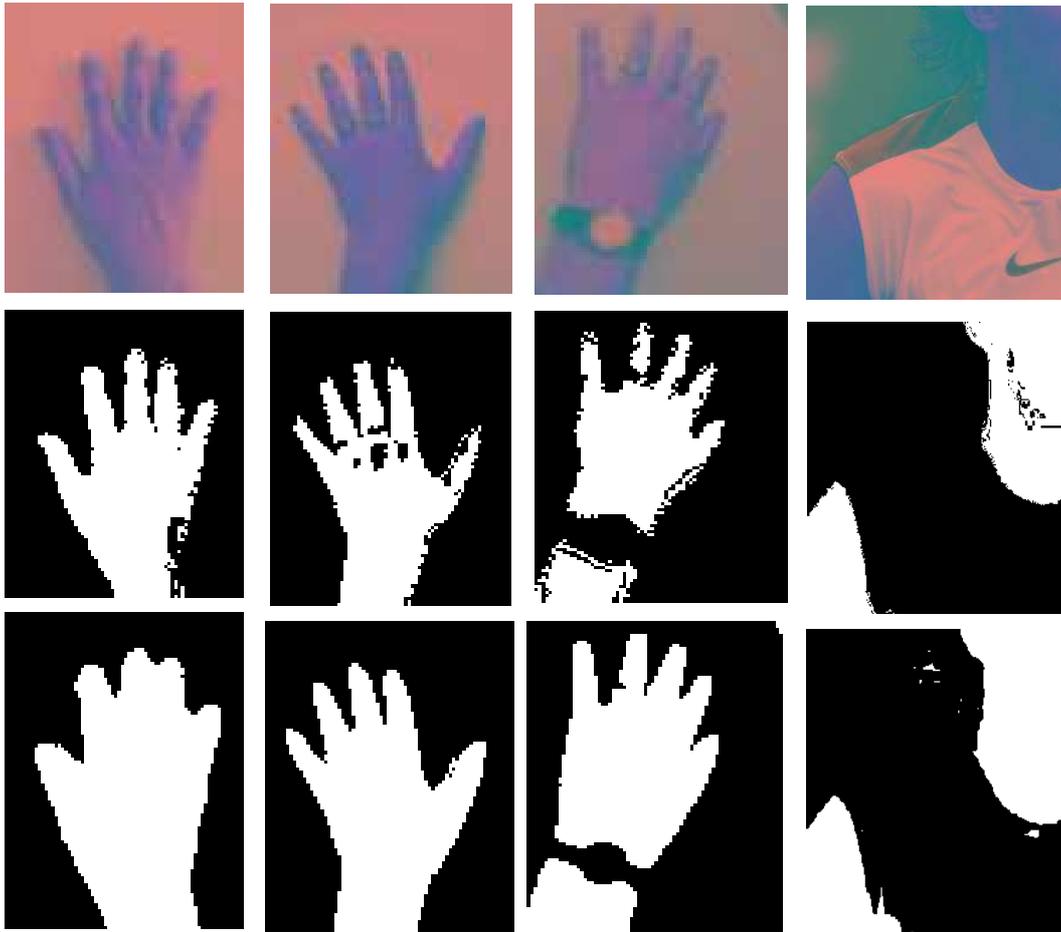


Fig.9 Détection de la peau à partir d'une image YCbCr la première ligne représente les images originales en YCbCr, la deuxième des images résultantes en utilisant un RBN et la troisième celles obtenues par des seuils de la couleur de la peau

Le tableau (3.2) représente les matrices de confusion, les taux de bonnes détections, les vrais positifs et les faux positifs obtenus en appliquant les deux méthodes de détection de la peau : le classifieur bayésien naïf et la méthode par seuillage dans des images YCbCr.

Tab.3.2 résultats obtenus en appliquant le RBN et les seuils pour la détection de la peau à partir des images YCbCr

	Matrice de confusion	Taux de bonnes détections	VP	FP
Réseau bayésien	$\begin{bmatrix} 4461 & 290 \\ 308 & 1852 \end{bmatrix}$	91.35%	85.78 %	13.43%
Seuillage	$\begin{bmatrix} 4451 & 300 \\ 617 & 1543 \end{bmatrix}$	86.73%	71.47 %	13.89%

La classification des pixels peau et non peau par un RBN dans une image YCbCr présente un taux de bonnes détections meilleur que celui de seuillage, avec des taux de bonnes détection 91.35% et 86.73%, vrais positifs de 85.78% et 71.47%, faux négatifs de 13.43% et 13.89% pour les méthodes du réseau bayésien et le seuillage respectivement. Ceci montre que pour une image YCbCr, le RBN est plus performant en terme de classification de pixels peau et non peau que la méthode basée sur des seuils fixés pour la détection de peau.

Résultats obtenus dans l'espace HSV

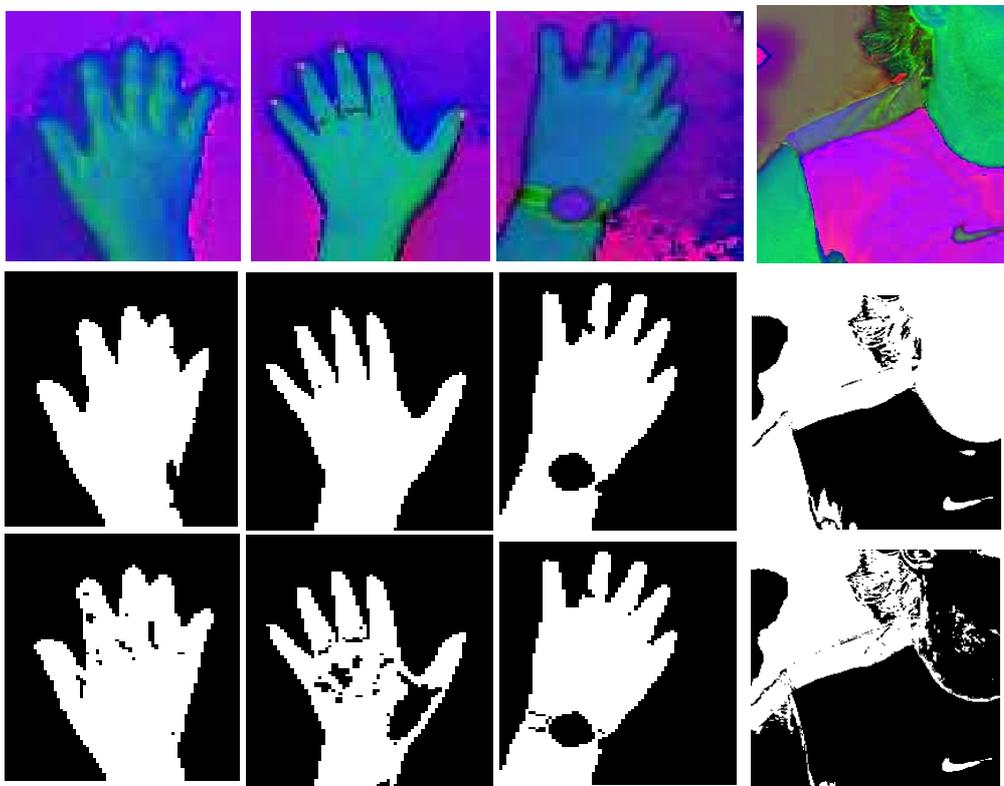


Fig. 10 Détection de la peau à partir d'une image HSV, la première ligne représente les images originales en HSV, la deuxième les images obtenues par un RBN et la troisième celles obtenues en appliquant la méthode des seuils de la couleur de la peau

Tab.3.3 Résultats obtenus en appliquant le RBN et les seuils pour la détection de la peau à partir des images HSV

	Matrice de confusion	Taux de détection	VP	FP
Réseau bayésien	$\begin{bmatrix} 4355 & 396 \\ 116 & 2044 \end{bmatrix}$	92.59%	94.67 %	18.34%
Seuillage	$\begin{bmatrix} 4413 & 338 \\ 466 & 1694 \end{bmatrix}$	88.36 %	78.46 %	15.65%

Dans ce dernier cas, où les images test sont en HSV le tableau (3.3) montre que le RBN donne des meilleurs résultats : 92.59% de taux de détection, 94.67% de taux de vrais positifs et 18.34 % de taux de faux positifs par rapport à 88.36%, 78.46% et 16% respectivement pour le seuillage. A partir de ces résultats, on peut conclure que l'espace HSV est opérant pour la détection de la peau et les réseaux bayésiens restent un outil de classification plus efficace que la méthode du seuillage.

4.2 Discussion des résultats

Pour le réseau bayésien naïf, les meilleurs taux sont obtenus dans les deux espaces YCbCr et HSV, cela est dû à l'élimination de la composante luminance (V pour l'espace HSV et Y pour YCbCr). Les fausses détections sont dues au mauvais éclairage et à certains objets qui prennent des couleurs semblables à celles de la peau humaine, chose que même l'œil humain ne peut distinguer dans certains cas. Dans l'espace RGB, les taux de détection sont moins bons, cela revient à la grande sensibilité de l'espace RGB à l'intensité d'éclairage (luminosité) à cause de la combinaison linéaire des valeurs de R, G et B qui représente la luminance, c'est-à-dire que le changement de luminance d'un patch de peau donné affecte toutes les composantes R, G et B. De plus, cet espace couleur ne sépare pas la luminance de la chrominance.

Nous remarquons que les résultats sont satisfaisants en utilisant les deux approches de classification, mais les réseaux bayésiens restent le moyen le plus performant grâce à l'opération d'apprentissage. Le désavantage de la méthode de seuillage persiste dans le choix des seuils de définition de la peau.

5. Conclusion

Nous avons évoqué dans ce chapitre les différentes méthodes de détection de la peau, nous nous sommes intéressées à la classification des pixels peau et des pixels non-peau qui est l'objet de notre travail. A partir de cette application, on peut dire que les réseaux bayésiens naïfs donnent des résultats probants, voire meilleurs que ceux de la méthode des seuils, ce qui avantage notre approche. Le choix de l'espace couleur est une étape primordiale dans la détection de la peau. Il varie selon la méthode utilisée ; dans notre cas les espaces couleurs les plus appropriés sont le HSV et le YCbCr avec des taux de bonnes détections de 92,59% et de 91,35%, du fait qu'ils distinguent la luminance de la chrominance.

Conclusion générale et perspectives

Ce mémoire est consacré à représenter un nouveau modèle graphique probabiliste dit *réseau bayésien* (BN) utilisé dans plusieurs domaines tels que le marketing, l'informatique, la médecine, l'industrie,..., etc. L'objectif de notre travail est d'introduire les réseaux bayésiens en traitement d'images dans un cadre de classification supervisée en utilisant un réseau bayésien spécifique appelé *réseau bayésien naïf* (NBN).

Ce travail nous a donc permis de nous familiariser à plusieurs domaines; à savoir le traitement d'images et plus particulièrement la classification supervisée, les réseaux bayésiens, la détection de la peau et la programmation sous MATLAB.

Au seuil de notre travail, nous avons abordé les notions sur le traitement d'images. Cependant, nous avons noté qu'un système de traitement d'images est composé de cinq étapes : acquisition, prétraitement, traitement numérique et enfin analyses et interprétations des images. La classification qui constitue l'objet de notre étude est une méthode de segmentation qui est un traitement d'images de bas niveau, c'est-à-dire nécessitant très peu d'informations sur le contenu des images. Les méthodes de classification se déclinent en deux catégories : classification supervisée et classification non supervisée. La première dispose d'un ensemble d'apprentissage et les classes doivent être connues a priori. Par contre, dans la deuxième, seul des exemples sans étiquettes sont disponibles. Après avoir évoqué les notions sur le traitement d'images, nous avons passé en revue les fondements théoriques des réseaux bayésiens où nous nous sommes intéressés aux méthodes d'apprentissage et d'inférence ainsi qu'aux extensions de ces derniers.

Les réseaux bayésiens sont actuellement l'une des méthodes les plus importantes de l'intelligence artificielle car ils permettent la représentation de la connaissance par un graphe acyclique orienté, compréhensible, dont les informations peuvent être incomplètes. De plus, comme ils sont basés sur des probabilités, ils intègrent la notion d'incertitude dans le raisonnement. La structure graphique des réseaux bayésiens permet le raisonnement de façon bidirectionnelle, autrement dit, suivant des relations

Conclusion générale et perspectives

de dépendances entre les variables dans les deux directions. Ce type de réseaux présente un inconvénient majeur qui est leur compréhension qui devient difficile dans le cas où l'on dispose d'un grand nombre de variables et/ou de liens de dépendances.

Afin de montrer l'efficacité des réseaux bayésiens, nous avons appliqué un réseau bayésien appelé *réseau bayésien naïf* (NBN) pour la détection de la peau en classifiant les pixels en pixels peau et pixels non peau dans des images couleur. Puis, nous avons comparé les résultats obtenus avec une méthode classique de détection de la peau basée sur des seuils fixés de définition de la couleur de la peau.

La détection de la peau dans des images couleur est une étape très délicate surtout parce qu'elle est considérée comme une étape préliminaire dans la plus part des systèmes homme-machine, détection de mouvement, détection, reconnaissance de visages,..., etc. Les performances de ces systèmes sont étroitement liées aux résultats obtenus dans cette étape. Cette tâche facile à réaliser par l'œil humain est restée une problématique ardue à résoudre automatiquement par un ordinateur. Ils existent plusieurs facteurs qui en sont à l'origine : la diversité des conditions de prise de vue, celle de la présence de plusieurs couleurs de la peau humaine ainsi que des objets ayant une couleur semblable à celle de la peau. Notre modèle de peau bayésien naïf a prouvé sa robustesse par rapport à toutes ces situations.

Dans le but d'obtenir une bonne classification des pixels peau et des pixels non peau, un choix des espaces couleur à utiliser s'avère d'une grande importance. Cela nous a amené à réaliser nos expériences sur trois espaces couleur (RGB, YCbCr et HSV), mais les résultats obtenus sont plus en faveur des espaces couleur perceptuels HSV et YCbCr (avec des taux de bonne détection de 92,59% et 91,37% pour les deux espaces respectifs) grâce à l'avantage qu'ils présentent : la distinction entre la luminance et la chrominance. La comparaison effectuée entre la méthode du seuillage et le réseau bayésien naïf nous permet de conclure que ce dernier est mieux adapté à la détection de la peau à condition de lui fournir une base d'apprentissage riche en couleur de peaux humaines.

Pour améliorer ce modeste travail et pour mieux prouver l'efficacité du classifieur bayésien naïf, il est important de citer quelques perspectives :

- Se servir d'une base de données riche utilisée généralement dans tous les travaux effectués sur la détection de la peau ; c'est la base « *Compaq* » de Jones et Rehg [68] .
- Utiliser un logiciel spécialement conçu pour les réseaux bayésiens appelé « *BayesiaLab* » qui représente un laboratoire complet de

Conclusion générale et perspectives

manipulations de réseaux bayésiens, il a été créée en 2001 par deux docteurs en sciences : Paul Munteanu et Lionel Joffe.

- Appliquer la détection de la peau dans un espace couleur propre à cette dernière introduisant l'information texture ; c'est l'espace « THS » développé par Forsyth et Fleck [69]
- Appliquer les réseaux bayésien à d'autres techniques de traitement d'image comme par exemple la détection et le suivi d'un objet en mouvement par les réseaux bayésiens dynamiques.

Références Bibliographiques

- [1] : **Steven W. Zucker. Region growing**, “Childhood and adolescence. Computer Graphics and Image Processing”, volume-5, N°3, pages 382–399, 1976.
- [2]: **Ludovic Macaire, Sylvie Philipp-Foliguet** : « Segmentation d’images couleur ».
- [3] : **Alexandre Hamon**, « Segmentation couleur de plans cadastraux de la ville de Paris au XIXème siècle ».
- [4] : **A. Boujelida**, « Réseaux Bayésiens Naifs augmentés TAN pour les systèmes de détection d’intrusions », Mémoire Magister en ingénierie des systèmes informatiques, Institut Nationale de formation en Informatique, Alger, 2008.
- [5] : : **J.Pearl**, “Probabilistic Reasoning in intelligent Systems”, Morgan Kaufmann, San Mateo, CA, 1988;
- [6]: **Ann Becker, Patrick Naim**, “Les réseaux bayésiens” edition EYROLLES, 1999.
- [7] : **Kh. HALLOULI**, « reconnaissance de caractères par Méthode Markoviennes et réseaux bayésiens », thèse doctorat, Ecole Nationale supérieure des Télécommunications, Paris, Mai 2004.
- [8] : **S . Baghdadi**, « Extraction multimodale de métadonnées de séquences vidéo dans un cadre bayésien », Thèse doctorat, Université de RENNE 1, Février 2010.
- [9] : **Jensen, F.V., Lauritzen, S.L., et Olesen, K.** “ Bayesian updating in causal probabilistic networks by local computations”, Comp.Stat.Quart, 4, 269-282. 1988.
- [10] **R.Detcher**, “Bucket elimination: a unifying framework for probabilistic inference”. In Proceeding of the NATO Advanced Study Institute on learning in graphical models, pages 75-104, 1998.
- [11] **J. H. Kim and J. Pearl.** “ A computational model for combining causal and diagnostic reasoning in inference systems”, In Proceedings of the Eighth International Joint Conference on Artificial Intelligence, Karlsruhe, Allemagne, 1983.
- [12] **S. Lauritzen & D. Spiegelhalter** “Local computations with probabilities on graphical structures and their application to expert systems”, Journal of the royal statistical society, series B, 50(2):157-224,1988.
- [13] **F. Jensen, S. Lauritzen & K. Olsen.** “Bayesian updating in causal probabilistic networks by local computations”. Computational Statistics Quarterly, 4:269-282,(1990).
- [14] **Zhang, N. & Poole, D.** (1994). A simple approach to bayesian network computations. Dans In Proceedings of the tenth Canadian Conference on Artificial Intelligence, (pp. 171–178). 44
- [15] **Shachter, R. D. (1986).** Intelligent probabilistic inference. Dans Kanal, L. N. & Lemmer, J. F. Eds.), Uncertainty in Artificial Intelligence, (pp. 371–382)., Amsterdam. North-Holland. 44

- [16] : **Patrick Naim, Pierre-Henri Wuillemin, Philippe Leray, Olivier Pourret, Anna Backer**, « Réseaux Bayésiens », édition EYROLLES , 2007.
- [17] E.Horvitz, J. Suermondt, et G.Cooper « Bounded conditioning: Flexible inference for decisions under scarce resources » In Proceedings of the fifth Conference on uncertainty in Artificial Intelligence, pages 182-193. North Holland, 1989.
- [18] **Bruce Dambrosio**. «Incremental probabilistic inference». In proceeding of the ninth Conference on Uncertainty in Artificial Intelligence, pages 301-308. Morgan Kaufman, 1993.
- [19] **P. Spirtes, C. Glymour and R. Scheines**. “Causation prediction and search”. New York, 1993, Springer Verlag.
- [20] **H. Akaike**. “Statistical predictor identification”. Ann. Inst. Statist. Math., 22 :203–217, 1970.
- [21] **G. Schwartz**. “Estimating the dimension of a model”. The Annals of Statistics, 6(2) :461–464, 1978.
- [22] **Pat Langley, Wayne Iba, and Kevin Thompson**. “An analysis of bayesian Classifiers”. In Proceedings of the Tenth National Conference on Artificial Intelligence, pages 223–228, San Jose, CA, 1992. AAAI Press.
- [23] **I. Tsamardinos, L.E. Brown et C.F. Aliferis**, “The max-min hill-climbing Bayesian network structure learning algorithm”, Machine Learning, vol. 65, no. 1, pages 31_78, 2006.
- [24] **J. Pearl**. “Causality: Models, Reasoning, and Inference”. Cambridge, England University Press, Cambridge, England, 2000.
- [25] **F.Olivier** « De l'identification de structure de réseaux bayésiens à la reconnaissance de forme à partir d'informations complètes ou incomplètes », Institut national des sciences appliquées de Rouen, 2006.
- [26] **D. Koller. & A. Jeffer**. (1997). “Object-oriented bayesian network”. Dans Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-97), (pp. 302–313).16
- [27] **O. Bangsø, & P. H. Wuillemin**. (2000). “Object oriented bayesian networks a framework for topdown specification of large bayesian networks and repetitive structures”. Technical Report CIT-87.2-00-obphw1, Hewlett-Packard Laboratory for Normative Systems, Aalborg University.16
- [28] **R. Howard, R. & J. Matheson**. (1981). “Influence diagrams”. In Howard, R. and Matheson, J., editors, Readings on the Principles and Applications of Decision Analysis, volume II, 721–762. 17
- [29] **R. Shachter, & C. Kenley**. (1989). “Gaussian influence diagrams”. Management Science, 35, 527–550. 17
- [30] **P. S. Maybeck**. (1979). “Stochastic models, estimation, and control”. volume 141 of Mathematics in Science and Engineering. Academic Press, Inc. (copyright

now owned by Navtech Seminars & GPS Supply).
http://www.cs.unc.edu/~welch/media/pdf/maybeck_ch1.pdf. 20

[31] **B. Anderson, & J. Moore,**. (1979). "Optimal Filtering. Prentice-Hall". 20

[32] **M. Takikawa, B. d'Ambrosio, & E. Wright,**. (2002). "Real-time inference with large scale temporal bayes nets". Dans Proceedings of the Seventeenth Conference of Uncertainty in Artificial Intelligence, (pp. 477–484)., San Mateo. Morgan Kaufmann. 18

[33] **J. Tian, & J. Pearl.** (2002a). "A general identification condition for causal effects" . Dans Proceedings of the Eighteenth National Conference on Artificial Intelligence, AAAI Press/The MIT Press :Menlo Park, CA, 567-573, August 2002, (pp. 567–573)., Menlo Park, CA. AAAI Press/The MIT Press. also printed in UCLA Cognitive Systems Laboratory Technical Report (R-290-A), April 2001. 24

[34] **U. Nodelman, & E. Horvitz.** (2003). "Ctbns for inferring users' presence and activities with extentions for modeling and evaluation". Technical Report MSR-TR-2003-97, Microsoft Research.21

[35] **K. Laskey.** (2006). "First-order bayesian logic". George Mason University Department of Systems Engineering and Operations Research.
http://ite.gmu.edu/~klaskey/papers/Laskey_FOBL.pdf. 21

[36] **L. A Zadeh,**. (1978). "Fuzzy sets as a basis for a theory of possibility". Fuzzy sets and systems, 1978,1, 3–28. 5, 25

[37] **S. Maes, S. Meganck, & B. Manderick.** (2005). "Identification of causal effects in multi-agent causal models". Dans Proceedings of the IASTED International Conference on Artificial Intelligence and Applications (AIA 2005), (pp. 178–182)., Innsbruck. 16, 24

[38] **S. Maes,**. (2005). "Multi-Agent Causal Models : Inference and Learning". PhD thesis, Vrije Universiteit Brussel, Faculteit Wetenschappen, DINF, Computational Modeling Lab. 24

[39] **S. Lauritzen. et D. Spiegelhalter,** (1988). « Local computations with probabilities on graphical structures and their application to expert systems". Royal statistical Society series B (Methodological), 50(2) :157–224.

[40] **U. Lerner, E. Segal et D. Koller,**. (2001). « Exact inference in networks with discrete children of continuous parents". In Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI-01), pages 319–32, San Francisco, CA. Morgan Kaufmann.

[41] : **S.Mitra and T.Acharya Recognition :** "A survey. IEEE Transactions on systems, Man and Cybernatics", Part C: Applications and Reviews. Volume 37, Issue.

[42]: **A. Hamamoorthy, N. Vswani, S. Chaudhury, and S. Benerjee.** "Recognition of dynamic handgesture".

- [43]: **A. Jacquin and A.E Ieftheriadis**, “Automatic location tracking of faces and facial features in video signal”, Int.work. On automatic face and Gesture recognition, pages: 142-147, Zurich, 1995.
- [44]: **K.Aas**, “Detection and Recognition faces in videosequences”, Norsk regnesentral, 1998.
- [45]: **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 226-227.
- [46]: **H. Zheng, M. Daoudi, and B. Jedynak**. “Blocking Adult Images Based on Statistical Skin Detection”. Electronic Letters on Computer Vision and Image Analysis. Volume 4, November2, 2004, Pages: 1-14.
- [47]: **L.Sigal, S.Sclaroff, and V. Athitsos**. “Skin colour – based video segmentation under time-varying illumination”. IEEE Transactions on Pattern Analysis and Machine Intelligence, Pages: 862-877, July 2004.
- [48]: **Gu. Lifang and D.Bone**. “Skin colour region detection in MPEG video sequences”. International Conference on Images Analysis and Processing. September 27-29 1999, Venice, Italy. Pages: 898-903.
- [49]: **M.Hammami**, « Modèle de peau et application à la classification d'image et au filtrage des sites web » Thèse doctorat, Ecole nationale de Lyon, 2005.
- [50]: **C.R. Wren, A.Azarbayjani, T.Darrell and A.Pentland**, “Pfinder: Real-time tracking of the human body”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7),pp.780-785,1997.
- [51]: **AHLBERG, J.** 1999. “A system for face localization and facial feature extraction”. Tech. Rep. LiTH-ISY-R-2172, Linkoping University.
- [52]: **CHAI, D., BOUZERDOUM, A.** 2000. “A bayesian approach to skin color classification in ycbcr color space”. In *Proceedings IEEE Region Ten Conference (TENCON'2000)*, vol. 2, 421-424
- [53]: **M.C. Shin, K.I. Chang, L.V. Tsap**, “Does colorspace transformation make any difference on skin detection?” In: WACV: Proceedings of the 6th IEEE Workshop on Applications of Computer Vision, Orlando, Florida, USA, 2002
- [54]: **P. Kakamanu, S. Makrogiannis, and N. Bourbakis.**” A survey of Skin-Color Modelling and Detection Methods”. Pattern Recognition. Volume 40, March 2007, Pages: 1106-1122.
- [55]: **V. Vezhnets, V. Sazonov.** “A survey on skin color detection techniques”, GRAPHICON03,2003,pp.85-92.
- [56]: **Z. Xu, and M. Zhu.** “Color- Based Skin Detection”. Twelfth Modelling Conference. January 4-6 2006. Pages: 143-152.

[57]: B. D. Zarit, B. J. Super, and F.K.H.Quek. "Comparaison of five color models in skin pixel classification". In Proceedings. International Workshops on Recognition, Analysis, and Trackings of Faces and Gestures in Real Times Systems, Pages 58-63, Corfu, Greece, September 1999.

[58]: J. C. Terrillon, M. N. Shirazi, H. Fukamachi, and S. Akanatsu. "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.

[59]: N. Friedman, D. Geiger, M. Goldszmidt, "Bayesian network classifier", Mach, Learn. 29(2) (1997) 131-163.

[60]: F. Bérard, « Vision par ordinateur pour l'interaction homme-machine fortement couplée », thèse de doctorat, Université Joseph Fourier, Grenoble, France, Janvier 2000.

[61]: F. K. H. Queck, T. Mysliwrec, and M. Zhao. "Fingermouse: A Freehand pointing inference". Proceedings of the International Workshop on Automatic Face and Gesture – Recognition, Zurich, Switzerland, pp.372-377, June 1995.

[62]: M. J. Jones, and J. M. Rehg. "Statistical color models with application to skin detection", International Journal of Computer Vision.

[63]: 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.

[64]: P. Peer, J. Kovac, and F. Solina, "Human skin colour clustering for face detection". International Conference on Computer as a Tool, EUROCON 2003, Ljubljana, Slovenia, September 2003.

[65]: L. Jordao, M. Perrone, J.P. Costeira, and J. Santos-Victor. "Active face and feature tracking". In Proceedings of International Conference on Image Analysis and Processing, pages 572-576, Venice, Italy, September 1999.

[66]: J. Ahlberg, "Extraction and Coding of Face Model Parameters", Licentiate Thesis No. 747, Department of Electrical Engineering, Linköping University, Sweden, March 1999.

[67]: V Vezhnevets, V Sazonovm, A Andreeva, "A Survey on Pixel-Based Skin Color Detection Techniques", *Proceedings on Graphicon*, 2003.

[68]: M. Jones and J. Rehg, "Statistical color models with application to skin detection", *IJCV*, 46(1):81-96, 2002.

[69]: M. M. Fleck, D. A. Forsyth, C. Bregler, "Finding naked people, Finding naked people", Proceedings of European Conference on Computer Vision, vol 2, 1996, pp. 592-602.

Annexes

Annexe A Théorie des graphes

Définition A.1 –Graphe

Soit $V = \{v_1, \dots, v_n\}$ un ensemble fini non vide. Un graphe G sur V est défini par la donnée du couple

$$G = (V, E) \text{ où } E \subset \{(u, v) \mid u, v \in V \text{ et } u \neq v\}$$

V est alors nommé l'ensemble des nœuds de G .

Définition A.2 -Arrête et Arc

Soit n graphe $G = (V, E)$, pour tout élément $(u, v) \in E$,

- (u, v) est une arrête (noté $(u - v)$) si et seulement si $(v, u) \in E$,
- (u, v) est un arc (noté $(u \rightarrow v)$) si et seulement si $(v, u) \notin E$.

Dans un arc, les deux éléments de V ne jouent pas le même rôle alors que dans une arrête, ces éléments sont symétriques.

Définition A.3 -Graphes orientés, non orientés, mixtes

Un graphe $G = (V, E)$ est un graphe orienté (noté \vec{G}) si et seulement si tous les éléments de E sont des arcs.

De même, G est un graphe non orienté (noté \bar{G}) si et seulement si tous les éléments de E sont des arêtes. Un graphe mixte est un graphe ni orienté, ni non orienté.

Définition A.4 –Parent, Enfant, racine, feuille

Soit un graphe orienté $G = (V, E)$, pour tout arc $(u \rightarrow v) \in E$, u est l'origine de l'arc, v est son extrémité, u est alors un parent (ou prédécesseur) de v ; v est l'enfant (ou successeur) de u .

Une racine d'un graphe est un nœud sans parent. Une feuille est un nœud sans enfant.

Définition A.5 –chemin, circuit

Dans un graphe orienté $G = (V, E)$, un chemin est une séquence d'arcs $(e_i)_{i \in \{1, \dots, p\}}$ vérifiant la propriété suivante : l'origine de tout arc e_i est l'extrémité de l'arc e_{i-1} précédent dans la séquence.

Annexes

Un circuit est un chemin dont l'extrémité du dernier arc est l'origine du premier.

Un chemin simple est un chemin dont lequel aucun arc n'apparaît plus d'une fois.

Définition A.6 –Chaine, cycle

Dans un graphe quelconque $G = (V,E)$, une chaîne est une séquence d'arêtes $(e_i)_{i \in \{1 \dots p\}}$ vérifiant la propriété suivante : pour tout $i \in \{2 \dots p-1\}$, l'une des extrémités d'une arête e_i est une extrémité de l'arête e_{i-1} précédente ; l'autre extrémité de e_i est une extrémité de l'arc suivant e_{i+1} .

Un cycle est une chaîne dont une extrémité du dernier arc est une extrémité du premier.

Définition A.7 -Connexion

Soit $G = (V,E)$ un graphe, soit u et v deux nœuds de G . on dit que u et v sont connectés si et seulement s'il existe un chemin $[u, v_0, v_1, \dots, v_n, u]$.

Définition A.8 -Graphe connecté

Soit $G = (V,E)$ un graphe, G est connecté si, entre deux nœuds quelconque, il existe au moins un chemin.

Définition A.9 -Graphe connecté acyclique orienté (CDAG)

Soit $G = (V,E)$ un graphe orienté. G est acyclique si G ne contient aucun cycle orienté.

Définition A.10 –Sous graphe, graphe complet, Clique

Soit un graphe $G = (V,E)$, $\forall W \subset V, \forall F \subset E, (W,E \cap W \times W)$ est un sous graphe de G .

Un graphe G est complet si et seulement si $\forall u, v \in V, u \neq v, (u \leftrightarrow v) \in E$; $(u \leftrightarrow v) \Leftrightarrow ((u \rightarrow v) \text{ ou } (v \rightarrow u) \text{ ou } (u - v))$.

Les cliques d'un graphe sont les sous-graphes complets maximaux (c'est-à-dire de cardinal maximal).

Définition A.11 Arbre, Arborescence

Un graphe $G = (V,E)$ est un arbre si et seulement s'il est connexe et sans cycle.

Un graphe G est une arborescence si et seulement si G est un arbre et possède une unique racine.

Annexe B Probabilités

Rappels de probabilités

Soient Ω un espace observable et \mathcal{A} une tribu d'évènements sur Ω . (Ω, \mathcal{A}) est un espace probabilisable.

Définition B.1 – Probabilité

Une application $P : \mathcal{A} \rightarrow [0, 1]$ est dite probabilité sur l'espace probabilisable (Ω, \mathcal{A}) si elle vérifie les axiomes suivants :

- $P(\Omega) = 1$
- Si $(A_n)_{n \in \mathbb{N}}$ est une famille d'évènements deux à deux disjoints, alors :

$$P(\cup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} P(A_n)$$

(Ω, P) est alors appelé espace probabilisé.

Définition B.2 – Variable aléatoire

Une variable aléatoire est une fonction X définie sur Ω :

$$X : \begin{cases} \Omega & \rightarrow D_x \\ w & \rightarrow X(w) \end{cases}$$

Pour $X \in D_x$, on note alors $\{X=x\}$ l'évènement $\{w \in \Omega \mid X(w) = x\}$.
 D_x est le domaine de définition de X .

Si le domaine de définition de la variable X est fini, alors X est une variable aléatoire discrète.

Pour la suite, la variable aléatoire sera représentée par une majuscule (X, Y, \dots). la valeur que prend cette variable aléatoire sera noté par la même lettre mais minuscule ($x \in D_x, y \in D_y, \dots$).

On simplifiera la notation de l'évènement représenté par $\{X = x\}$, de telle façon que :
 $P(\{X=x\}) = P(X=x) = p(x)$.

Définition B.3 – Probabilité jointe

Soit A et B deux variables aléatoires sur le même univers Ω . On parle alors de probabilité pour la fonction définie sur D_A et D_B par :

$$P_{AB} : \begin{cases} D_A \times D_B & \rightarrow [0,1] \\ (a, b) & \rightarrow P_{AB}(a, b) = P(\{A = a\} \cap \{B = b\}) \end{cases}$$

Définition B.4 – marginalisation

Annexes

Soit U un ensemble fini, non vide de variables aléatoires, $V \subset U$ non vide et $V' = U \setminus V$ et $P(U)$ la probabilité jointe sur les variables de U , on appelle alors marginalisation de P sur V la fonction :

$$\forall v \in D_V, P(v) = \sum_{v' \in D_{V'}} P(v, v')$$

Définition B.5 -Espérance, variance, écart-type

Nous appellerons espérance de la variable X , la valeur $E(X) = \sum_X X \cdot P(X)$, sa variance, la valeur $V(X) = E((X-E(X))^2)$, et son écart-type, le nombre $\sigma(X) = \sqrt{V(X)}$.

Définition B .6 (vraisemblance et log-vraisemblance)

Soit X un vecteur de variables aléatoires suivant une distribution multinomiale d'ordre 1 de fonction de densité P .

La fonction de vraisemblance de la table de probabilité P fondé sur N observations de X est égale à la probabilité d'obtenir cet ensemble d'observations dans l'ordre.

$$L(P, N) = \prod_{x \in X} P(X)^{N(X)}$$

La log-vraisemblance de la table de probabilités P est égale à l'opposé de logarithme de cette fonction de vraisemblance, pour obtenir une quantité positive.

$$L(P, N) = - \sum_{x \in X} N(X) \log(P(X))$$

Définition B.10 (distribution de Dirichlet)

Soient $\alpha_1, \alpha_2, \dots, \alpha_r$ tous strictement positifs. Une distribution de Dirichlet d'exposants $\alpha_1, \alpha_2, \dots, \alpha_r$ possède une densité de probabilité de la forme suivante :

$$\text{Dir}(\theta \mid \alpha_1, \alpha_2, \dots, \alpha_r) = \frac{\Gamma(\sum_{i=1}^r \alpha_i)}{\prod_{i=1}^r \Gamma(\alpha_i)} \prod_{i=1}^r \theta^{\alpha_i - 1}$$

Où Γ est la fonction Gamma qui satisfait $\Gamma(x+1) = x\Gamma(x)$ et $\Gamma(1) = 1$

Hypothèse de Dirichlet : les densités de probabilités des paramètres admettent des densités exponentielles de la forme de Dirichlet.

Annexe C Bayes net toolbox (BNT)

Présentation

La Bayes Net Toolbox (BNT) est une bibliothèque open-source de fonction Matlab pour la création, l'inférence et l'apprentissage de modèle graphique dirigés ou non dirigés, disponible sur <http://bnt.sourceforge.net>. Elle a été programmée par Keven Murphy en 1997, et bénéficie maintenant du soutien de nombreux chercheurs qui y apportent de nouvelles fonctions régulièrement, faisant de BNT un outil précieux pour tous les chercheurs.

Elle s'avère être un très bon outil pour l'utilisation des réseaux bayésiens car :

- elle effectue l'inférence bayésienne soit de manière exacte, soit de manière approchée, avec plusieurs moteurs implémentés ;
- elle gère également les distributions de probabilités discrètes, gaussiennes et les Ou-bruités binaires ;
- elle traite l'apprentissage des paramètres ;
- elle gère l'apprentissage de la structure avec IC*, PC et K2.

De plus cet outil récupère les avantages de Matlab, qui a de bons algorithmes numériques, et les tableaux sont particulièrement bien adaptés à la gestion des tables de probabilités conditionnelles des réseaux bayésiens.

Apprentissage

➤ Paramètres

BNT est capable d'estimer les paramètres d'un réseau bayésien à partir de données complètes (par maximum de vraisemblance ou maximum a posteriori) ou de données incomplètes grâce à l'algorithme EM.

➤ Structure

Concernant l'apprentissage de structure, BNT met à notre disposition plusieurs fonctions de score comme BIC ou le critère BDe. Les algorithmes K2, IC/PC et IC*/PC* sont aussi disponibles.

Inférence

BNT prend en charge plusieurs algorithmes d'inférence, ainsi pour des réseaux bayésiens discrets, gaussiens ou mixtes (conditionnels gaussien) :

- Elimination de variable .
- Arbre de jonction.
- Algorithme de pearl exact (pour les polyarbres) ou approché.
- Par échantillonnage : likelihood weighting et Gibbs sampling.