

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
DEPARTEMENT D'AUTOMATIQUE

**Mémoire de Fin d'Etudes
de MASTER ACADEMIQUE**

Domaine : **Sciences et Technologies**
Filière : **Génie électrique**
Spécialité : **Commande des systèmes**

Présentée par

DIHA Sadia

LEKADIR Amina

Encadré par

DIAF Moussa

Thème

***Réseaux Bayésiens appliqués a la reconnaissance
automatique d'objets colorés***

Mémoire soutenu publiquement le / / 2015 devant le jury composé de

Année universitaire 2014-2015

Liste des abréviations

3D	Trois dimension
KPPV	K plus proches voisins
SVM	Les séparateurs à vaste marge
RGB	Red Green Blue
MGS	Modèle gaussian simple
GMM	Mélange de modèles gaussians
N	Nombres d'observations
K	Nombres des classes
C	Classes
$P(X/C_k)$	Lois de probabilités conditionnelles
$P(C_k)$	Désigne la probabilité a priori
$d_k(Xq)$	Les fonctions de décision
μ_k	La moyenne
$\sum k$	La variance
$V_k(x)$	Les points d'apprentissage
G	Graphe Acyclique orienté
V	L'ensemble des nœuds
E	L'ensemble des arcs
Ω	L'univers de probabilités
P	L'ensembles des variables aléatoires
X	Variable aléatoire
$P(A)$	probabilité a priori de A
Ns	nombres de structure

Liste des figures

Chapitre I : La reconnaissance faciale

Fig.1 : Architecture générique d'un système de reconnaissance biométrique.

Fig.2 : Difficultés de la reconnaissance de visage :(a)changements d'échelle. (b) changements d'éclairage. (c)port de lunettes. (d) changements d'orientations du visage. (e) changements des expressions faciales.

Chapitre II : Les réseaux Bayésiens

Fig.1 : Représentation graphique de la causalité.

Fig.2 : Sous-ensemble d'un arbre.

Fig.3 : Exemple d'un réseau bayésiens naïf.

Fig.4 : Réseau bayésien naïf augmenté par un arbre.

Chapitre III : Détection de la peau du visage

Fig.1. Exemple des images d'apprentissage

Fig.2. L'architecture de RBN avec les paramètres d'espace couleur HSV

Fig.3. L'architecture de RBN avec les paramètres de texture

Fig.4. L'architecture de RBN avec les paramètres de la texture et de couleur

Fig.5. Détection de la peau à partir de sa couleur : (a) Les images HSV, (b) Les images résultantes.

Fig.6. Détection de la peau à partir des paramètres de la texture. (a) Les images au niveau de gris, (b) Les images résultantes.

Fig.7. Détection de la peau en fusion les deux approche. (a) Les images tests originales, (b) Les images résultantes.

Fig.8. Images test : (a). Image original (b). Image au niveau de gris (c). Image en HSV (d), (e), (f) Les images résultantes par ordre (texture, couleur, la fusion (texture+couleur)).

Sommaire

Liste des abréviations

Liste des figures

Introduction générale 1

Chapitre I : La reconnaissance faciale

1.	Introduction.....	3
2.	Généralités sur les systèmes biométriques.....	3
3.	Les techniques biométriques.....	6
4.	La place de la reconnaissance faciale parmi les autres techniques biométriques	7
	4.1.Expression faciale	9
	4.2.Classification des expressions faciales	9
	4.2.1. Méthode supervisée	10
	4.2.2. Méthodes non supervisée	12
	4.3.Détection de visages	12
	4.3.1. Caractéristiques d'un détecteur de visage	13
	4.3.2. Les techniques de détection de visages	13
	4.4.Extraction des caractéristiques de visage	14
	4.5.La reconnaissance de visage	15
5.	Principales difficultés de la reconnaissance de visage	16
6.	Conclusion	17

Chapitre II : Les réseaux bayésiens

1.	Introduction	18
2.	Définition	19
3.	Formule de Bayes	20
4.	La décision Bayésienne	22
5.	Construction des réseaux bayésiens	23
6.	Principe du Réseau Bayésien	23
6.1.	Apprentissage dans les réseaux bayésiens	23
6.1.1.	Apprentissage de la structure	23
6.1.2.	Apprentissage des paramètres	24
6.2.	Inférence	26
6.2.1.	Algorithmes d'inférence exacte	26
6.2.2.	Algorithmes d'inférence approximative	30
7.	Réseaux bayésiens adaptés à la classification.....	30
8.	Conclusion	33

Chapitre III : Détection de la peau du visage

1.	Introduction.....	34
2.	Détection de la peau	34
2.1.	Approche basée sur la couleur	35
2.2.	Approche basée sur la texture.....	36
3.	La matrice de cooccurrence.....	36
4.	Classification par les réseaux bayésiens naïfs.....	36
5.	Illustration et interprétation des résultats	37
5.1.	Illustration des résultats	40
5.2.	Interprétation des résultats	45
6.	Conclusion	46
	Conclusion générale	47

Introduction générale

Le développement des technologies modernes a atteint ces dernières années des proportions jamais égalées du fait de l'amélioration notable des systèmes et processus biométriques. Divers domaines sont concernés par cette quête et en particulier la défense et la sécurité d'individu où le progrès est plus en plus rapide.

Parmi toutes les technologies biométriques qui existent, la reconnaissance faciale qui est une modalité qui puisse être implémentée de manière indépendante des autres modalités biométriques. Elle est souvent utilisée dans des applications de surveillance, contrôle d'accès, gestion horaire, paiement sécurisé sur Internet...etc.

Plusieurs méthodes sont développées pour mettre en œuvre cette technologie de reconnaissance faciale. La plus pratique est celle qui se base sur les modèles graphiques probabilistes, et plus précisément les réseaux bayésiens, qui sont révélés des outils très efficaces pour la représentation de connaissances incertaines, et le raisonnement à partir d'informations incomplètes. Le formalisme des réseaux bayésiens est donc l'un des rares outils scientifiques qui servent à la fois à prévoir et à détecter les objets, et cela, même lorsque la situation est incertaine et les données incomplètes. Les RB sont donc particulièrement adaptés dans plusieurs domaines d'application : on trouve le domaine de la santé, l'industrie, la défense, l'informatique et la sécurité.

Le présent travail se vaut plus au domaine de la sécurité dans lequel nous intéresserons à l'identification et la détection des personnes. Pour y a faire plusieurs techniques sont développées. Parmi elles, nous allons utiliser les techniques basées sur la détection du visage à partir de la couleur et la texture de la peau du visage. Notre choix du fait que l'information couleur et texture restent toujours les moyens les plus couramment utilisés par plusieurs chercheurs [53] [54] [55]. Cette technique peut se faire en deux étapes successives. La première consiste à caractériser chaque pixel par un ensemble d'attributs, la deuxième consiste à regrouper les pixels ayant les mêmes caractéristiques dans une même classe.

Ainsi, nous allons utiliser 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.

Le mémoire est principalement scindé en trois chapitres.

Dans le premier chapitre, nous exposerons d'abord les aspects fondamentaux relatifs à la biométrie et en particuliers à la reconnaissance faciale. Ensuite, nous allons étudier les principales composantes d'un système de reconnaissance faciale, notamment la détection, l'extraction des caractéristiques et la reconnaissance du visage. Enfin, nous mettrons en lumière les difficultés rencontrées par les systèmes de reconnaissance du visage.

Dans le deuxième chapitre, nous présenterons un bref aperçu sur les réseaux bayésiens. Nous allons décrire quelques algorithmes d'apprentissage et d'inférence les plus utilisés. Pour clore ce chapitre, nous allons établir quelques domaines d'utilisation des réseaux bayésiens ainsi leurs principaux avantages.

Le troisième chapitre est consacré à la description de la méthode adoptée pour la segmentation d'images couleur. Cette technique consiste à caractériser chaque pixel par des attributs de couleur et de texture en se basant sur le classifieur bayésien naïf et les matrices de cooccurrence. Enfin, nous allons fusionner adaptativement des informations de couleur et de texture. Les résultats de la classification obtenus par cette méthode seront également présentés.

Ce travail sera terminé par une conclusion générale dans laquelle nous résumerons nos contributions et présenterons une synthèse des résultats obtenus, et des références bibliographiques.

Chapitre I

La reconnaissance faciale

1. Introduction

La sécurité des systèmes d'information est devenue un domaine de recherche d'une très grande importance, en particulier, concevoir un système d'identification fiable, efficace et robuste est une tâche prioritaire. L'identification de l'individu est devenue essentielle pour assurer la sécurité des systèmes, face à cette sollicitation grandissante, plusieurs méthodes de reconnaissance biométriques ont été proposées : reconnaissance faciale, empreinte digitale, reconnaissance de l'iris, la forme de la main...etc.

Dans ce présent chapitre, nous exposerons la présentation générale des systèmes biométriques, puis nous allons définir les outils utilisés pour évaluer leurs performances. Ensuite, nous montrons la place de la reconnaissance faciale parmi les autres techniques biométriques. Puis à travers de ce chapitre, Nous allons étudier également les principales composantes d'un système de reconnaissance faciale, notamment la détection, l'extraction de caractéristiques et la reconnaissance. Enfin, nous mettrons en lumière les difficultés rencontrées par les systèmes de reconnaissance de visages.

2. Généralités sur les systèmes biométriques [1] [2] [3]

La biométrie recouvre l'ensemble des procédés tendant à identifier des êtres vivants, plus précisément, la biométrie est l'étude mathématique des variations biologiques à l'intérieur d'un groupe déterminé. Autrement dit, la biométrie est l'exploitation automatisée de caractéristiques physiologiques (empreintes digitales, visages, iris, contour de la main, ... etc) ou comportementales (signature, démarche, ... etc). Pour déterminer ou vérifier l'identité d'un individu, il existe traditionnellement deux manières :La première méthode est basée sur une connaissance à priori "knowledge-based" de la personne telle que, la connaissance de son code PIN qui permet d'activer un téléphone portable. La seconde

méthode est basée sur la possession d'un objet "token-based", il peut s'agir d'une pièce d'identité, d'une clef, d'un badge, ... etc. Cependant, chacune de ces deux méthodes ont leurs inconvénients. Dans le premier cas, le mot de passe par exemple peut être oublié par son utilisateur et dans le second cas, le badge, la pièce d'identité ou la clef peut être perdue ou volée. Pour que ces caractéristiques puissent être utilisées comme des biométries, elles doivent être universelles, uniques, permanentes, collectables et mesurables. L'intérêt majeur des ces applications biométriques se résumant en trois classes: faciliter le mode de vie, éviter la fraude et contrôler la population.

Un système biométrique est essentiellement un système de reconnaissance de formes qui utilise les données biométriques d'un individu. Il peut fonctionner en mode d'inscription ou en mode de vérification ou bien en mode d'identification.

- *Le mode d'inscription* est une phase d'apprentissage qui a pour but de recueillir des informations biométriques sur les personnes à identifier. Pendant cette phase, les caractéristiques biométriques des individus sont saisies par un capteur biométrique, puis représentées sous forme numérique (signatures), et enfin stockées dans la base de données.
- *Le mode de vérification* effectue une comparaison un à un pour valider l'identité d'une personne en comparant les données biométriques saisies avec le modèle biométrique de cette personne stocké dans la base de données du système.
- *Le mode d'identification* reconnaît un individu par une recherche sur l'ensemble de la base des modèles. Il effectue une comparaison un à plusieurs pour établir l'identité de l'individu. Ce mode consiste à associer une identité à une personne.

Il existe au moins deux modules dans un système biométrique : le module d'apprentissage et celui de reconnaissance le troisième module facultatif est le module d'adaptation. Pendant l'apprentissage, le système acquerra une ou plusieurs mesures biométriques qui serviront à construire un modèle de l'individu. Ce modèle de référence servira de point de comparaison lors de la reconnaissance. Le modèle pourra être réévalué après chaque utilisation grâce au module d'adaptation [4]. La figure d'après montre l'architecture générique d'un système de reconnaissance biométrique.

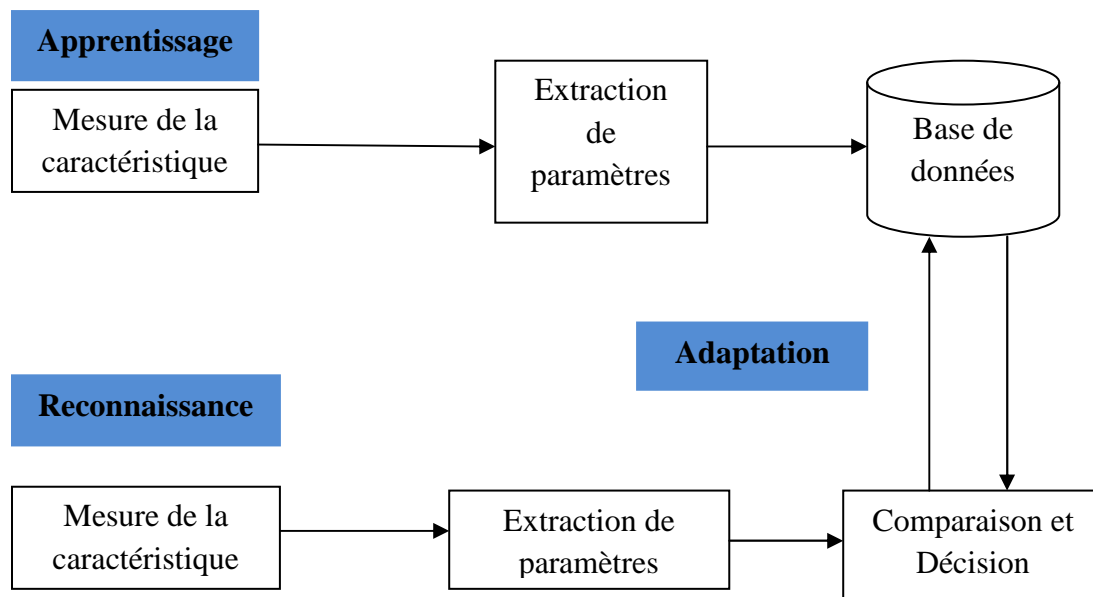


Fig.1. Architecture générique d'un système de reconnaissance biométrique

Dans la phase d'apprentissage, on mesure tout d'abord la caractéristique biométrique à l'aide d'un capteur. Le modèle obtenu par ce dernier est une représentation compacte du signal qui permet de faciliter la phase de reconnaissance, mais aussi de diminuer la quantité de données à stocker. Ce modèle peut être stocké dans une base de données représenté sur (Fig.1) ou sur une carte à puce. De même, lors de la phase de reconnaissance la caractéristique biométrique est mesurée. Le capteur utilisé doit avoir des propriétés aussi proches que possibles du capteur utilisé durant la phase d'apprentissage. Si les deux capteurs ont des propriétés trop différentes, il faudra en général appliquer une série de prétraitements supplémentaires pour limiter la dégradation des performances. La suite de la reconnaissance sera différente suivant le mode opératoire du système : identification ou vérification.

Pour construire un modèle assez général capable de décrire toutes les variations possibles de cet attribut. Le module d'adaptation peut se faire en mode supervisé ou non supervisé, il est donc nécessaire pour maintenir voire amélioré la performance d'un système utilisation après utilisation.

3. Les techniques biométriques [5] [6] [7]

Aujourd'hui, la biométrie est un domaine de recherche en pleine expansion, plusieurs systèmes d'identification et de vérification sont à présent développés. Cependant leurs performances restent insuffisantes face aux besoins accrus de plus de sécurité. L'utilisation d'une seule modalité biométrique diminue, dans la plupart des cas, la fiabilité de ces systèmes, ce qui nous a incités à combiner plusieurs modalités [8]. Nous ne donnerons ici qu'un rapide aperçu des technologies biométriques les plus courantes : Pour les caractéristiques physiques, nous décrirons la reconnaissance de visages qui est la plus communément employée par l'homme pour effectuer une identification personnelle. Par ailleurs, La reconnaissance à partir du visage se base sur les caractéristiques jugées significatives comme l'écart entre les yeux, la forme de la bouche, le tour du visage, la position des oreilles, la couleur de la peau,..., etc. Parmi les avantages de cette technique : non couteuse, peu encombrante, absence de contact avec le capteur, méthode non intrusive pour la personne et pas de risques pour la santé. De même, La reconnaissance d'empreintes digitales [9] qui est la technique biométrique la plus ancienne. Cependant les empreintes digitales sont une mesure biométrique assez mal acceptées par les utilisateurs à cause de l'association qui est souvent faite avec la criminologie. Les empreintes digitales sont formées par les crêtes (ridge) et les vallées (furrow) présentées sur la surface du bout des doigts. La reconnaissance de l'iris qui est une technologie la plus récente puisqu'elle ne s'est véritablement développée que dans les années 80. Cette technologie représente 6% des applications. Elles sont très intéressantes pour la biométrie, car elle permet de différencier même des jumeaux ou l'œil gauche de l'œil droit d'une même personne. En effet, La technologie de l'iris présente aussi une quasi-infinité de points caractéristiques qui ne varient pratiquement pas pendant la vie contrairement à sa couleur qui peut changer. Mais cela n'a aucune influence car les images d'iris obtenues par les capteurs sont en noir et blanc.

Pour les caractéristiques comportementales, nous décrirons les systèmes basés sur la voix, qui est la seule technique qui permette à l'heure actuelle de reconnaître une personne à distance. Cependant cette technique est très facilement falsifiable en utilisant un enregistrement. De plus, la méthode basée sur la signature dont on trouve chaque personne possède une signature qui lui est propre et qui peut donc servir à l'identifier. Il existe deux modes de reconnaissance : le mode statique et le mode dynamique. Le mode statique n'utilise que l'information géométrique de la signature. Le mode dynamique utilise à la fois

l'information géométrique et dynamique, c'est-à-dire les mesures de vitesse, d'accélération,... etc. Le mode dynamique est plus riche en information que le mode statique, il est donc plus discriminant.

Il existe d'autres méthodes biométriques basées sur les veines de la main, l'A.D.N. (acide désoxyribonucléique), l'odeur corporelle, la forme de l'oreille, la forme des lèvres, le rythme de frappe sur un clavier, la démarche,... etc.

Le choix de ces technique dépend de plusieurs facteurs [10] sont :

- La précision et la fiabilité de l'identification ou de la vérification. Le résultat ne doit surtout pas être susceptible aux conditions ambiantes ou au vieillissement.
- Le coût d'installation, du maintien et de l'opération.
- L'universalité de l'authentification par exemple, la reconnaissance par l'écriture ne convient pas aux analphabètes.
- La facilité d'emploi.
- La reproductibilité des résultats. En général, les caractéristiques physiologiques sont plus reproductibles et moins susceptibles à la contrefaçon que les caractéristiques comportementales.

4. La place de la reconnaissance faciale parmi les autres techniques biométriques

La biométrie est une science qui s'est développée depuis le début du siècle passé. Elle a été employée au début pour identifier les criminels. Cette technique initiée par A. Bertillon en 1903 s'appuyait sur des mesures anthropométriques du corps caractérisant les personnes. L'empreinte digitale est la technique biométrique la plus ancienne et la plus répandue et qui a été adoptée par Scotland Yard dès 1901 et par la Préfecture de Police de Paris, dès 1903 pour rechercher les criminels. La reconnaissance faciale, quant à elle, a trouvé des applications dans différents domaines comme l'identification des personnes inconnues et cela depuis 1926. L'expression faciale, déjà étudiée par Darwin et Duchenne de Boulogne au dix-huitième siècle, a joué un rôle majeur dans la recherche sur les émotions depuis les travaux de Sylvan Tomkins dans les années 1960.

De plus, l'utilisation des techniques de reconnaissance faciale a connu un développement à grande échelle depuis le milieu des années 90, avec l'utilisation efficace de nouvelles technologies, notamment l'ordinateur et sa capacité de traitement d'images et de

fouille de banques de données (datamining) ce qui lui a donné des nouvelles applications. Aujourd'hui la reconnaissance faciale est utilisée dans plusieurs domaines, telles que la sécurité, contrôle d'accès, ...etc. D'après [11], en 2010 après avoir mis à l'essai divers systèmes de reconnaissance faciale, le National Institute of Standards and Technology des États-Unis a constaté que le meilleur algorithme permettait de reconnaître avec exactitude 92 % des inconnus au moyen d'une base de données de 1,6 million de dossiers criminels.

Par ailleurs, aux États-Unis, cette technologie continue à se développer, elle est même utilisée par les agences gouvernementales (FBI) ou par des entreprises privées [12]. De même, la technologie de reconnaissance faciale fait de plus en plus partie intégrante des mesures de sécurité déployées dans le cadre d'événements sportifs de grande envergure. Lors des Jeux Olympiques de Pékin, en 2008, toutes les personnes qui entraient dans le stade principal devaient se soumettre à une vérification de l'identité à un poste muni d'un dispositif de reconnaissance faciale. On a également eu recours à cette technologie pour les jeux Olympiques de Londres en 2012 afin de surveiller l'entrée de suspects identifiés. Aussi lors de la Coupe du monde de 2014 [13], les services de police brésiliens prévoient utiliser des lunettes de style « Robocop » équipées d'une caméra capable de filmer 400 images par seconde et les comparées avec une base de données numérique de 13 millions de photos.

D'après un rapport publié en 2013 par le service de recherche du Congrès américain [14], les organismes chargés de l'application de la loi pourraient chercher prochainement à équiper des véhicules aériens sans pilote de dispositifs de reconnaissance faciale ou d'autres technologies de reconnaissance biométrique douces permettant de reconnaître et de surveiller des individus sur la base de caractéristiques telles que la taille, l'âge, le sexe et la couleur de la peau...etc.

En effet, la nouveauté dans la reconnaissance faciale arrive grâce au développement des nouvelles caméras de type 3D. Ces caméras obtiennent de meilleurs résultats que les caméras classiques, parce qu'elles acquièrent une image tridimensionnelle de chaque visage.

4.1. Expression faciale

L'expression faciale est un aspect important du comportement et de la communication non-verbale [15] où le changement dans le visage. En général, trois étapes principales peuvent être distinguées dans un système d'analyse d'expression faciale. La première étape consiste à détecter le visage, c'est l'étape qui permet de limiter la zone de recherche. Par la suite,

l'extraction des informations nécessaires qui décrivent au mieux l'expression. A la fin, on se basant sur ces informations, l'image sera affectée à une catégorie d'expressions à l'aide d'un classifieur.

Pour que la technologie soit efficace, il faut disposer d'une image numérique du visage de l'individu, d'une base de données d'images numériques d'individus identifiés et d'un logiciel de reconnaissance faciale capable d'établir une correspondance exacte entre l'image d'un individu et une image d'un individu identifié qui est enregistrée dans la base de données.

4.2. Classification des expressions faciales

La classification, d'une manière générale, permet de partitionner un ensemble d'observations en différentes classes, en regroupant dans une même classe les observations présentant les mêmes caractéristiques. Par ailleurs, la classification des images intervient comme une étape très importante dans des applications biométrie et la reconnaissance des formes telles que la reconnaissance des visages, de la parole, la reconnaissance par géométrie de la main, la détection de la peau, ... etc.

Les méthodes de classification sont divisées en deux groupes [16] [17] selon l'information disponible sur les spécifications de classes ; lorsque les différentes classes des observations sont connues au préalable pour la classification on parle de classification supervisée. Cette catégorie de méthodes est préférable pour la classification des images. En revanche, dans le cas où l'on n'a aucune connaissance a priori sur l'appartenance des observations aux classes, on parle de classification non-supervisée ou clustering. Celle-ci est plus adaptée pour la segmentation, elle ne nécessite aucune intervention humaine.

Nous présenterons ici un bref aperçu des méthodes de classification supervisée et non supervisée les plus courantes, ainsi leurs avantages et inconvénients respectifs.

4.2.1. Méthode supervisée

Dans le cadre de la reconnaissance de formes, les méthodes de classification supervisée tentent d'apprendre. À partir d'images pour lesquelles la classe est connue, constituant un échantillon d'apprentissage et une fonction de classification. Cette fonction permettra d'associer une valeur de classe à chaque image dans la classe inconnue.

Il existe une multitude de méthodes de classification supervisée, parmi les méthodes les plus classiques : la règle des K plus proches voisins(KPPV) [19] et les séparateurs à vaste

marge(SVM) [20], la classification optimale de Bayes [25], Dans notre travail nous s'intéressons à cet dernier.

Classification par la méthode Bayésienne

Cette méthode de nature statistique, on suppose que l'échantillon à analyser est composé de N observations tirées, indépendamment, de K classes C_k où $k=1,2,\dots, k$ selon les lois de probabilités conditionnelles $P(X/C_k)$ connues. Par application du théorème de Bayes, il est possible de déterminer la probabilité pour qu'une observation X_q appartienne à la classe C_k . Cette probabilité, appelée aussi probabilité a posteriori de la classe C_k , elle est donnée par la relation :

$$P(C_k|X_q) = \frac{p(X_q|C_k)p(C_k)}{p(X_q)} \quad (1)$$

Où $P(X_q)$ désigne la valeur de la fonction de densité sous-jacente à la distribution des observations provenant de l'échantillon au point X_q et $P(C_k)$ désigne la probabilité a priori de la classe C_k de telle sorte que :

$$\sum_{k=1}^K P(C_k) = 1 \quad (2)$$

$P(X_q/C_k)$ désigne la probabilité conditionnelle de la classe C_k , c'est-à-dire, la fonction de densité de probabilité sous-jacente à la distribution des observations provenant la classe C_k .

Le problème de classification d'une observation X_q peut être résolu en adoptant la règle de décision bayésienne suivante :

$$P(C_s/X_q) > P(C_k/X_q), \forall k=1, 2... K, (k \neq s) \quad (3)$$

L'application de la règle de Bayes revient à trouver des fonctions de décision $d_k(X_q)$, $k=1, 2, \dots, k$, telles que l'observation X_q soit affectée à la classe C_s si et seulement si :

$$d_s(X_q) > d_k(X_q) \quad \forall k=1,2,\dots,k, s \neq k \quad (4)$$

Dans la plus part des cas, la probabilité conditionnelle de la classe C_k est de type gaussienne telle que :

$$P(X_q|C_k) = \frac{\exp(-\frac{1}{2}(X_q|\mu_k))^T \Sigma_k^{-1}(X_q|\mu_k)}{(2\pi)^{\frac{N}{2}} |\Sigma_k|^{-\frac{1}{2}}} \quad (5)$$

Dans cette expression, μ_k est le vecteur moyen de la C_k et Σ_k la matrice de covariance de la classe C_k .

Ainsi, les fonctions de décision d_k peuvent être écrites sous la forme :

$$d_k(X_q) = -\frac{1}{2} \log|\Sigma_k| - \frac{1}{2} (X_q - \mu_k)^T \Sigma_k^{-1} (X_q - \mu_k) + \log[P(C_k)] \quad (6)$$

Les probabilités $P(C_k)$ peuvent être estimées par :

$$P(C_k) = \frac{N_{CK}}{N} \quad (7)$$

Où N_{CK} est le nombre d'observations appartenant à la classe C_k et N le nombre d'observations totales. μ_k et Σ_k sont estimés par la moyenne et la variance empirique à partir des prototypes de la classe C_k .

Les méthodes bayésiennes ont l'avantage d'être simples et efficaces, malgré les hypothèses d'indépendance entre les variables. De même, des études comparatives des algorithmes de classification ont prouvé que le plus simple classificateur Bayésien, appelé « naïve Bayes », avait des performances similaires à celles des arbres de décision et des réseaux de neurones.

De plus, ils se sont montrés précis et rapides appliqués à de grandes bases de données. Les méthodes de classification probabilistes, de ce fait, ont été largement utilisées en reconnaissance de formes, et continuent de l'être.

Par contre la classification Bayésienne est associée à certains désavantages par exemple, son application nécessite des probabilités dont la détermination requiert typiquement de grandes quantités de données ou plusieurs connaissances a priori. De plus les méthodes Bayésiennes nécessitent un coût de calcul relativement élevé pour déterminer l'hypothèse optimale dans un cas général. Encore ce modèle probabiliste n'est parfois pas un concept intuitif pour un expert du domaine.

4.2.2. Méthodes non supervisées

La classification non supervisée ou "clustering" est l'une des techniques fondamentales de l'extraction de données structurées ou non structurées. Plusieurs méthodes ont été proposées: Parmi celle l'ACP (l'Analyse en Composantes Principales, Principal Component Analysis) qui fait l'apprentissage en utilisant un nombre d'attributs de

données moins grand que le nombre d'attributs initiaux. Il y a aussi l'algorithme des K means qui est aussi connue sous le nom de méthode des centres mobiles. Cet algorithme simple, rapide mais reste toujours sensible au choix de la valeur k. Le but de l'algorithme K-means est de regrouper les données en k partitions différentes, le nombre de groupes k est défini par l'utilisateur. [18]

4.3. Détection de visages

La détection du visage a été très largement abordée par la communauté du traitement d'images et de la vision par ordinateur. Ainsi la reconnaissance d'identité, d'émotions et d'expressions, sont aujourd'hui des activités de recherche très actives. De plus, La détection des visages ou la localisation des visages dans une image d'entrée est une étape fondamentale de tout système de traitement des visages.

Plusieurs méthodes ont été vues présentement dans ce domaine, l'une des ces méthodes présentées par Yang et Waibel [24] consiste à détecter le visage en utilisant principalement la couleur du visage. La méthode de Yang et Waibel consiste à convertir l'image de l'espace colorimétrique RGB vers l'espace chromatique RG et ensuite de suivre un modèle de couleur qui représente le visage dans l'espace converti.

L'identification est souvent la finalité des applications de sécurité publique et nationale. Par exemple, un aéroport ou un centre sportif. La reconnaissance faciale convient très bien pour les applications d'identification, car les images du visage peuvent être captées à distance et à l'insu de l'individu. On peut aussi avoir recours à d'autres technologies biométriques, comme la reconnaissance de la démarche ou de la voix, pour identifier les individus à distance et sans leur consentement, mais elles comportent des limites évidentes qui les rendent moins utiles.

4.3.1. Caractéristiques d'un détecteur de visage

Un détecteur de visage idéal doit posséder les caractéristiques suivantes:

- *Robustesse*: Il doit être capable de gérer les variations d'apparence selon la pose, la taille, l'éclairage, l'occlusion, les milieux complexes, les expressions faciales et la résolution.
- *Rapidité*: Il doit être assez rapide pour effectuer le traitement en temps réel.

- *Simplicité*: Il doit être simple. Par exemple, le temps de formation doit être court, le nombre de paramètres doit être petit, et les échantillons d'apprentissage doivent pouvoir être recueillis à moindre coût.

4.3.2. Les techniques de détection de visages

Il existe une variété de techniques consacrées à la détection de visage. Les systèmes de suivi et de détection du visage se sont beaucoup développés ces dernières années grâce à l'amélioration du matériel et à la forte demande industrielle, par exemple pour la recherche d'individu par le biais d'une caméra de surveillance, ou encore l'utilisation de robots d'exploration (recherche de survivants à la suite d'un incendie). Une variété technique consacrée dans ce domaine, Parmi lesquelles, les méthodes basées sur les caractéristiques invariantes « Feature invariant approaches » [26].

Les algorithmes contenus dans cette catégorie se basent sur le fait que les caractéristiques faciales ne changent pas même quand l'orientation du visage, l'angle de prise de vue ou les conditions d'éclairage varient. On trouve ainsi dans cette catégorie des techniques basées sur la texture [27] ou la signature de couleur de la peau [25] pour la détection des différentes caractéristiques faciales. Tel que, la couleur et la texture constituent les deux sources d'information les plus utilisées dans plusieurs applications d'analyse d'images, la reconnaissance d'objets, l'indexation multimédia, la segmentation d'image, etc.

- Détection de visages basée sur l'analyse de la couleur de la peau

Les méthodes de détection basées sur l'analyse de la couleur de la peau [23] sont des méthodes efficaces et rapides. Elles réduisent l'espace de recherche de la région visage dans l'image. De plus, la couleur de la peau est une information robuste face aux rotations, aux changements d'échelle, et aux occultations partielles. Les techniques de détection du visage basées sur la couleur de la peau peuvent être classifiées en quatre catégories que nous allons maintenant aborder: les méthodes explicites qui consiste à définir explicitement les frontières de la région peau dans l'espace couleur utilisé. Tel qu'elles utilisent des règles de décision empiriques et/ou statistiques pour la détection des pixels ayant la couleur de la peau, les méthodes non paramétriques, ces approche ne dépendent pas de la forme de la fonction de distribution de la teinte. Elles utilisent les histogrammes couleur 2D et 3D pour représenter la densité de probabilité de la tonalité de la peau dans un espace couleur, de plus, leurs performances dépendent directement des images d'apprentissage.

Le besoin d'un modèle plus compact pour représenter la peau, et qui peut être généralisé avec moins de données d'apprentissage, a poussé les chercheurs à développer des modèles paramétriques de distribution de peau. Cette dernière est souvent représentée sous la forme d'un modèle gaussien simple (MGS) parfois elliptique, ou par un mélange de modèles gaussiens (GMM) et les méthodes semi paramétriques qui est principalement utilisées pour classifier des données de grande dimension, mais elles fonctionnent tout aussi bien pour des données de faible dimension, leur objectif est de classifier les pixels d'entrée selon qu'ils correspondent ou non à des pixels de peau, ce qui permet au final d'extraire la région du visage. Toutes ces approches pratiquent une phase d'apprentissage sur un nombre d'images représentatives pour calculer une densité de probabilité de la couleur peau.[2]

➤ Détection de visages basée sur la texture

La texture est un ensemble de pixels (niveaux de gris) spatialement agencés selon un certain nombre de relations spatiales, ainsi créant une région homogène. Il ya plusieurs approches et modèles qui ont été proposées pour la caractérisation de la texture, parmi les plus connue on a la matrice de cooccurrence. Qui est largement utilisée dans le monde du traitement d'image. Elle présente une grande simplicité de mise en œuvre et donne de bons résultats sur la plupart des types d'images.

4.4. Extraction des caractéristiques de visage

En plus de la classification, l'étape de l'extraction des paramètres représente le cœur du système de reconnaissance, elle consiste à effectuer le traitement de l'image dans un autre espace de travail plus simple et qui assure une meilleure exploitation de données, et donc permettre l'utilisation, seulement, des informations utiles, discriminantes et non redondantes. Dans le but de trouver un petit nombre de caractéristiques qui, une fois combinées, forment un classifieur efficace. On fait appliquer la méthode d'apprentissage, qui consiste à mémoriser les paramètres, après extraction et classification, dans une base de données bien ordonnées pour faciliter la phase de reconnaissance et la prise d'une décision.

4.5. La reconnaissance de visage

La reconnaissance du visage est la technique la plus commune et la plus populaire parmi les techniques de reconnaissance biométrique. Ainsi, la reconnaissance du visage est utilisée dans un grand nombre d'applications incluant la sécurité, le contrôle d'accès, le fichage, la communication et le loisir informatique.

Les systèmes de reconnaissance automatique des expressions du visage peuvent être utilisés principalement pour l'interaction homme-machine. Bien que beaucoup de progrès ont été accomplis, la reconnaissance de l'expression du visage, en temps réel, avec une grande précision reste un défi en raison de la complexité et la variabilité de ces expressions.

De nombreuses méthodes de reconnaissance de visage ont été proposées durant les trois dernières décennies[22]. De ce fait, la littérature sur cette thématique est très vaste et en aborde des aspects très variés. Les méthodes de reconnaissance de visage peuvent être regroupées suivant qu'elles opèrent une caractérisation globale, locale ou hybride des visages, ce que nous définissons dans ce qui suit :

Méthodes globales

Ces méthodes utilisent la totalité du visage comme entrée du système de reconnaissance. Une des représentations les plus utilisées pour coder le visage est basée sur les Eigenfaces, qui s'appuie sur l'analyse en composantes principales. De plus, elles sont basées sur des techniques d'analyse statistique bien connues. Dans ces méthodes, les images de visage (qui peuvent être vues comme des matrices de valeurs de pixels) sont traitées de manière globale et sont généralement transformées en vecteurs, plus faciles à manipuler. L'avantage principal des méthodes globales est qu'elles sont relativement rapides à mettre en œuvre et que les calculs de base sont d'une complexité moyenne. En revanche, elles sont très sensibles aux variations d'éclairage, de pose et d'expression faciale. Ceci se comprend aisément puisque la moindre variation des conditions de l'environnement entraîne des changements inéluctables dans les valeurs des pixels qui sont traités directement.

Méthodes locales

Dans ces méthodes, les caractéristiques locales telles que les yeux, le nez et la bouche sont extraites et leurs positions ainsi que leurs statistiques locales (issues d'informations géométriques, colorimétriques ou texturales) sont fournies à un classifieur structurel. Toutes ces méthodes ont l'avantage de pouvoir modéliser plus facilement les variations de pose, d'éclairage et d'expression par rapport aux méthodes globales. Toutefois, elles sont plus lourdes à utiliser puisqu'il faut souvent placer manuellement un assez grand

nombre de points sur le visage alors que les méthodes globales ne nécessitent de connaître que la position des yeux afin de normaliser les images, ce qui peut être fait automatiquement et de manière assez fiable par un algorithme de détection.

Méthodes hybrides

Les méthodes hybrides permettent d'associer les avantages des méthodes globales et locales en combinant la détection de caractéristiques géométriques (ou structurales) avec l'extraction de caractéristiques d'apparence locales. Elles permettent d'augmenter la stabilité de la performance de reconnaissance lors de changements de pose, d'éclairage et d'expressions faciales. Ces méthodes peuvent potentiellement obtenir de meilleurs taux de reconnaissance que les deux types de méthodes précédentes.

5. Principales difficultés de la reconnaissance de visage

La reconnaissance de visage est la méthode la plus naturelle car c'est la méthode utilisée inconsciemment par tous les individus grâce à l'interaction visuelle. Elle est surtout utilisée dans des casinos, des stades, mais également dans les aéroports.

Le problème de ces méthodes est le manque d'efficacité si les images en entrée sont de faible luminosité ou si certaines parties d'un visage sont masquées. De plus, le fait qu'un visage soit ombragé rend l'extraction des caractéristiques faciales très difficiles.

Cependant, cette technique fait face à beaucoup de difficultés en raison de la très grande variabilité de la forme à détecter (image d'un visage quelconque, d'orientation et de taille quelconque avec un éclairage quelconque), les états d'expression faciales, d'occlusion et d'éclairage changent également l'aspect global des visages, faible luminosité, mais aussi le port de lunettes ou autre modification visible du visage (opération, blessure, pilosité, maquillage,...). De plus, si certaines parties d'un visage sont masquées, et le fait qu'un visage soit ombragé rend l'extraction des caractéristiques faciales très difficiles, et moins d'efficacité (fig.2).

En fait, il y a pas mal en détection de visage dus aux conditions rigoureuses pour la classification des formes et les techniques d'apprentissage[21].

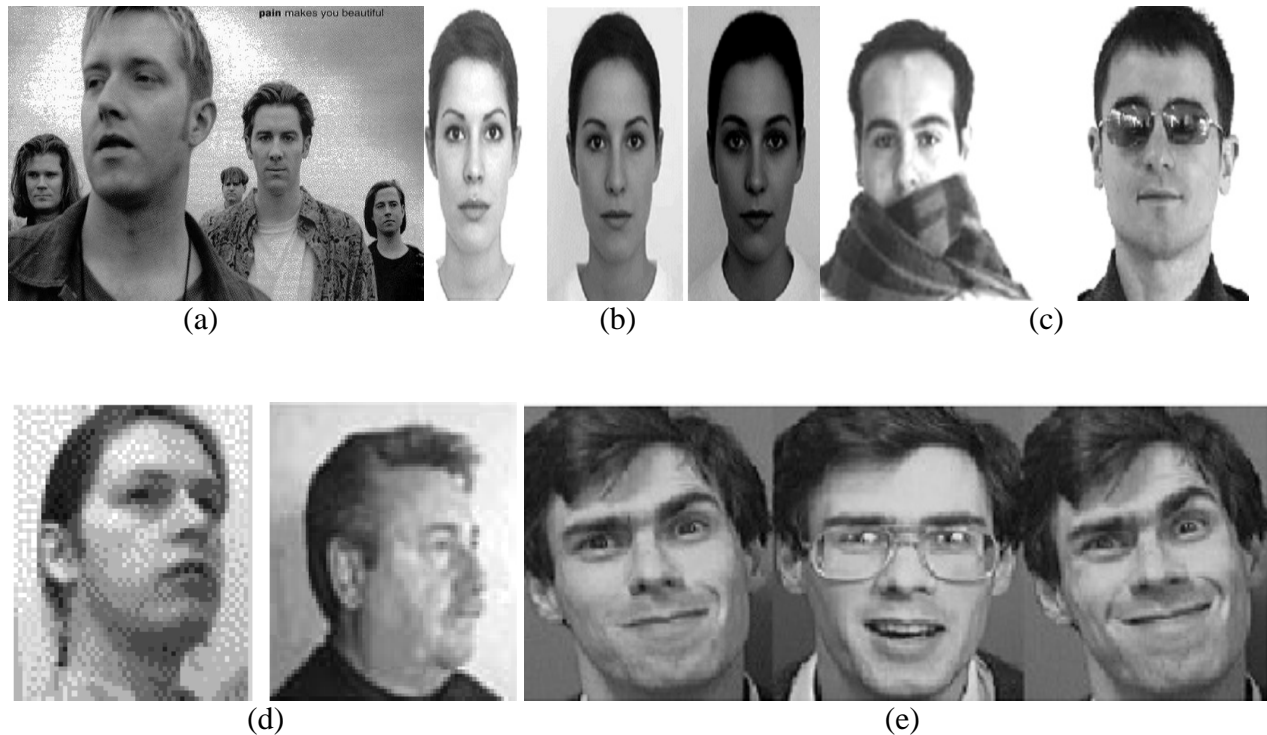


Fig.2. difficultés de la reconnaissance de visage : (a)changements d'échelle. (b) changements d'éclairage. (c)port de lunettes. (d) changements d'orientations du visage. (e) changements des expressions faciales.

6. Conclusion

Dans ce chapitre nous avons mis en relief quelques notions et définitions de base liées à la biométrie et ses diverses technologies. Après une comparaison des technologies biométriques on a choisit la reconnaissance faciale pour sa popularité et sa simplicité ainsi le coût faible de réaliser ce système parce qu'il suffit un caméra pour capter les images de visages et un micro-ordinateur pour faire les calculs. Où on a cité quelque méthodes de classification utilisé dans ce domaine, dont nous s'intéresserons à la classification bayésien car c'est l'objectif du deuxième chapitre. Enfin, nous avons mis en évidence les différentes difficultés inhérentes à la reconnaissance automatique de visages. Cette étude nous a permis de constater que la reconnaissance de visage suscite de plus en plus l'intérêt de la communauté scientifique, car elle présente plusieurs challenges et verrous technologiques.

Chapitre II

Les réseaux Bayésiens

1. Introduction

Les réseaux bayésiens sont le résultat des recherches effectuées dans les années 1980, dus à l'informaticien américain Judea Pearl(1936) à UCLA (University of California, Los Angeles) et à une équipe de chercheurs danoise à l'université d'Aalborg. L'objectif initial de ces travaux était d'intégrer la notion d'incertitude dans les systèmes experts. Les RB est une union entre la théorie des probabilités et la théorie des graphes afin de donner des outils efficaces pour représenter une distribution de probabilités jointes sur un ensemble de variables aléatoires. Leur utilisation nous permettre d'obtenir une représentation compacte de ces ensembles de dépendances grâce à la notion de d-séparation et des tables de probabilités conditionnelles (**Annexe. B**), à partir de laquelle il lui sera plus simple de raisonner. Les RB permettent aussi de modéliser et de capitaliser des connaissances (connaissances d'experts, exploitation de bases de données, de détection de fraude, reconnaissance des formes ...), d'établir des diagnostics (applications en médecine et en maintenance), de modéliser et d'aider à la prise de décision.

Dans ce chapitre nous allons définir les réseaux bayésiens. Ainsi nous présenterons les principales définitions, les principaux théorèmes et les étapes de la construction des réseaux bayésiens. Ensuite, nous parlerons de quelques algorithmes d'inférence et d'apprentissage, enfin nous étudierons le réseau bayésien adapté à la classification (réseau bayésien naïf), et quelques avantages liées à ce dernier.

2. Définition

les RB sont des modèles graphiques probabilistes représentant la connaissance et permettant de calculer des probabilités conditionnelles en s'appuyant sur le théorème de Bayes qui est détaillé dans le paragraphe suivant. Les RB apportent des solutions à différentes sortes de problématiques [28]. Ils décrivent la distribution des probabilités associées à un ensemble de variables dont certains sont directement dépendantes et d'autres sont indépendantes conditionnellement [29]. L'intérêt particulier de ces réseaux est de tenir compte simultanément de connaissances a priori d'experts et de l'expérience contenue dans les données. [30]

Autrement dit, un réseau bayésien est un système représente la connaissance, il permet de calculer des probabilités conditionnelles apportant des solutions à différentes sortes de problématiques. De plus, les réseaux Bayésiens sont un moyen simple de représenter une distribution de probabilité jointe d'un ensemble de variables aléatoires et de visualiser les propriétés de dépendance conditionnelle. Leurs dépendances sont représentées par ses arcs et codées en tant que probabilités de distributions conditionnelles aux niveaux des nœuds. Tel que, si une variable aléatoire X dépend de Y, alors le nœud X a un arc directionnel du nœud Y.

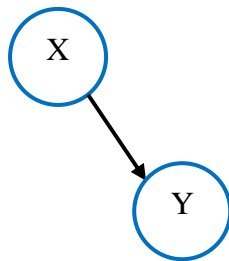


Fig.1. Représentation graphique de la causalité

De même, 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 jouent un rôle de plus en plus important dans la conception et l'analyse d'algorithmes d'apprentissage par machine (machine learning algorithms). Les principaux avantages s'inscrivent dans la description du modèle (la structure d'un modèle graphique montre de manière simple et claire les dépendances conditionnelles dans le modèle de

probabilités correspondantes, permettant l'évaluation et la révision du modèle), et dans l'efficacité du calcul.

Définition Formelle

D'une façon formelle, un réseau bayésien est constitué de la réunion d'un graphe acyclique orienté G , tel que : $G = G(V, E)$ où V est l'ensemble des nœuds de G , E l'ensemble des arcs de G et d'un espace probabilisé fini (Ω, p) , avec Ω est l'univers de probabilités et p l'ensemble des variables aléatoires $X = \{X_1, \dots, X_n\}$ associées aux nœuds du graphe tel que :

La distribution de la probabilité jointe $P(X)$ sur un ensemble de variables aléatoires

$X = \{X_1, \dots, X_n\}$ de ce modèle, s'écrit de la façon suivante :

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i / Pa(X_i)) \quad (1)$$

Où $Pa(X_i)$ est l'ensemble des parents du nœud X_i dans G .

3. Formule de Bayes

Le théorème de Bayes est un résultat de base en théorie des probabilités, issu des travaux du révérend Thomas Bayes (1702-1761), présenté en 1763. Il est utilisé dans l'inférence statistique pour mettre à jour les estimations d'une probabilité ou d'un paramètre quelconque à partir des observations et des lois de probabilités.

Pour aboutir au théorème de Bayes, on part d'une des définitions de la probabilité conditionnelle :

$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A) \quad (2)$$

En notant $P(A \cap B)$ la probabilité que A et B aient tous les deux lieu. En divisant de part et d'autre par $P(B)$, on obtient :

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

Le terme $P(A)$ est la probabilité a priori de A . elle est aussi appelée la probabilité marginale de A . Le terme $P(A|B)$ est appelée la probabilité a posteriori de A sachant B (ou encore de A sous condition B). Elle est « postérieure », au sens qu'elle dépend directement de B . Le terme $P(B|A)$ pour un B connu, est appelé la fonction de vraisemblance de A . De même, le terme $P(B)$ est appelé la probabilité marginale ou a priori de B .

On améliore parfois le théorème de Bayes [31] en remarquant que :

$$P(B) = P(A \cap B) + P(\bar{A} \cap B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) \quad (4)$$

Afin de réécrire le théorème ainsi :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})} \quad (5)$$

Où \bar{A} est le complémentaire de A. Plus généralement, si $\{A_i\}$ est une partition de l'ensemble des possibles, on a :

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)} \quad (6)$$

Autres écritures du théorème de Bayes

➤ Théorème de Bayes généralisé (règle de chaînage)

La distribution de probabilité jointe globale associée à l'ensemble des variables aléatoires est donnée selon le théorème de Bayes par :

$$\begin{aligned} P(X_1, \dots, X_n) &= P(X_1) P(X_2|X_1) P(X_3|X_1, X_2) \dots P(X_n|X_1, X_2, \dots, X_{n-1}) \\ &= P(X_1) \prod_{i=2}^n P(X_i|X_{i-1}, \dots, X_1) \end{aligned} \quad (7)$$

Mais dans un réseau bayésien, les probabilités conditionnelles de certaines variables X_i ne sont pas dépendantes de tous les prédécesseurs de X_i (c-à-d X_1, \dots, X_{i-1}), mais seulement 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 ou simplement les parents notés $Pa(X_i)$, alors nous pouvons écrire :

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

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

$$P(x_1, \dots, x_n) = \prod P(x_i/\text{parents}(x_i)) \quad (9)$$

Cette formule est également appelée formule de factorisation ou de chaînage.[32]

4. La décision Bayésienne

C'est une méthode de classification automatique optimale de la reconnaissance des formes formulée comme suit :

Un 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_k . Cette probabilité appelée aussi probabilité a posteriori de la classe C_k , elle est donnée par la relation:

$$P(C_r / X) = \frac{P(X/C_r)P(C_r)}{p(X_r)} \quad (10)$$

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

$$P(X/C_j) P(C_j) \geq P(X/C_k) P(C_k) \quad k \neq j, 1 \leq r \leq k \quad (11)$$

5. Construction des réseaux bayésiens

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.

La construction d'un réseau bayésien se décompose en trois étapes distinctes citées au dessous :

- Identification des variables et de leurs espaces d'états : cette étape dite qualitative consiste à la définition de l'ensemble des variables du système, avec précision de l'espace d'état de chaque variable.
- Définition de la structure du réseau bayésien est l'étape probabiliste introduit l'idée de distribution jointe définie sur les variables à générés de la base d'observations et choisir une structure de graphe qui sera compatible avec les variables.
- Définition de la loi de probabilité conjointe des variables est l'étape quantitative consiste en l'évaluation numérique des distributions de probabilités conditionnelles.

6. Principe du Réseau Bayésien

Les RB permettent d'effectuer des calculs complexes comme l'apprentissage des probabilités et l'inférence avec des manipulations graphiques.

L'aspect graphique de ces modèles est le fait de pouvoir construire ce graphe à l'aide d'expertise et/ou d'algorithmes d'apprentissage font les RB des outils de plus en plus utilisés en extraction et en gestion des connaissances [33]. Ces réseaux cherchent à fournir à cette base une modélisation sous forme de graphe caractérisant les dépendances conditionnelles des différentes variables. Leurs études nécessitent une base de données qui se déroule en deux phases: Inférence et Apprentissage.

6.1.Apprentissage dans les réseaux bayésiens

L'apprentissage d'un réseau bayésien sachant un ensemble d'observations peut se décomposer en deux étapes. La première consiste à apprendre sa structure, c'est-à-dire le graphe associé au réseau. Une fois sa structure est connue, il est nécessaire d'estimer les paramètres du réseau correspondant aux distributions de probabilités conditionnelles des variables.

6.1.1. Apprentissage de la structure

L'apprentissage de la structure des RB est un problème NP-difficile, qui a mené à la proposition de nombreuses méthodes. Ces dernières sont basées sur des données d'observation qui peuvent être classées sommairement en trois grandes familles. Les méthodes basées sur des tests d'indépendance conditionnelle, les méthodes à base de score et les méthodes hybrides qui sont une combinaison entre les deux méthodes précédentes.

En 1977, Robinson a montré que le nombre de structure N_s générés possible à partir de n nœud est donné par la formule suivante :

$$N_s(n) = \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} N_s(n-i), \quad n > 1 \quad (12)$$

Ce qui donne pour : $n=4$ on a $N_s(4)=543$ et pour : $n=7$ on a $N_s(7)=1.4 * 10^9$.

Comme l'équation (12) est super-exponentielle, il est impossible d'effectuer un parcours exhaustif en un temps raisonnable dès que le nombre de nœuds dépasse 7 ou 8, ce qui rend les méthodes d'apprentissage de la structure sont applicables seulement pour les populations à

taille limitée (quelques centaines de variables). De raison, ces méthodes utilisent alors une heuristique de recherche dans l'espace des graphes acycliques dirigés (DAG).

➤ *Les méthodes à base de recherche d'indépendances conditionnelles*

C'est la première série d'approches, consiste à rechercher les différentes indépendances conditionnelles qui existent entre deux variables quelconques conditionnellement à un ensemble de variables, dans le but de réduire l'espace de recherche de la meilleure structure. Cette série proposée initialement par une équipe de chercheurs Spirtes, Glymour, Scheines en 1993[34] d'un côté, avec les algorithmes SGS, P C, CI, FCI et d'autre côté, Pearl et Verma avec les algorithmes IC et IC*[35]. Le principe de ces algorithmes est basé sur les étapes suivantes :

- 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.

➤ *Les méthodes à base de score*

Contrairement à la première famille de méthodes, les deuxièmes approches vont soit chercher la structure qui va maximiser un certain score, soit chercher les meilleures structures et combiner leurs résultats. Pour que ces approches à base de score soient réalisables en pratique. Il faut les décomposé localement. C'est-à-dire s'exprimer comme la somme de scores locaux au niveau de chaque nœud ou un produit de mesures dont chacune n'est fonction seulement que d'un nœud et de l'ensemble de ses parents :

Exemple, si on prend un RB où $B = (G, \theta)$ 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 :

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

➤ *Les méthodes hybrides*

Les méthodes hybrides sont dans le but de combiner les avantages des méthodes basées sur la recherche d'indépendance et celles basées sur les scores. Ce rapprochement peut s'effectuer en deux étapes. Une recherche locale et Une optimisation globale. La recherche locale, permet d'obtenir un voisinage contenant toutes les dépendances

ou indépendances locales intéressantes avec l'aide de tests d'indépendances. L'optimisation globale, permet de faire une recherche sur l'espace des DAG en se restreignant aux dépendances locales trouvées précédemment.

6.1.2. Apprentissage des paramètres

Après avoir trouvé la structure du réseau bayésien pendant l'apprentissage de structure, il est nécessaire d'estimer les distributions de probabilités conditionnelles du réseau. Comme pour tout problème d'apprentissage, différentes techniques sont possibles selon la disponibilité de données pour le problème à traiter, ou d'experts du domaine. On peut classer ces techniques en deux grandes familles : apprentissage à partir de données complètes ou non, et par des approches statistiques classiques ou bayésiennes.

À partir de données complètes

Dans le cas où toutes les variables sont observées, la méthode la plus simple et la plus utilisée est l'estimation statistique qui consiste à estimer la probabilité d'un événement par la fréquence d'apparition de l'événement dans la base de données. Cette approche appelée maximum de vraisemblance (MV) [36], nous donne alors :

$$P(X_i = x_k \mid \text{Pa}(X_i) = x_j) = \theta_{i,j,k}^{MV} = \frac{N_{i,j,k}}{\sum_k N_{i,j,k}} \quad (13)$$

Où $N_{i,j,k}$ est le nombre d'événements 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 .

Le principe est quelque peu différent de l'estimation bayésienne qui consiste à trouver les paramètres θ les plus probables sachant que les données ont été observées. En utilisant une distribution de Dirichlet comme a priori sur les paramètres, on peut écrire :

$$P(X_i = x_k \mid \text{Pa}(X_i) = x_j) = \theta_{i,j,k}^{MAP} = \frac{N_{i,j,k} + \alpha_{i,j,k} - 1}{\sum_k (N_{i,j,k} + \alpha_{i,j,k} - 1)} \quad (14)$$

Il existe aussi, l'approche bayésienne qui consiste à considérer l'espérance à la place du maximum. Il s'agit alors, de l'espérance a posteriori qui nous permet de calculer ces paramètres par :

$$P(X_i = x_k \mid \text{Pa}(X_i) = x_j) = \theta_{i,j,k}^{EAP} = \frac{N_{i,j,k} + \alpha_{i,j,k}}{\sum_k (N_{i,j,k} + \alpha_{i,j,k})} \quad (15)$$

À partir des données incomplètes

Dans le cas des données incomplètes, Il existe plusieurs algorithmes pour estimer les données manquantes à partir des données connues au lieu de les ignorer. Parmi ces algorithmes, MCAR (Missing Completely At Random), MAR (Missing At Random), NMAR (Not Missing At Random) et les algorithmes EM (Expectation-Maximisation ou Espérance maximisation). Ces derniers sont le plus utilisés [36] car ils sont les plus faciles à résoudre. Ils sont composés de quatre étapes itératifs généraux d'estimation des paramètres θ qui peuvent être utilisés dans les deux cas d'apprentissage statistique et d'apprentissage bayésien. Le principe de ces algorithmes est présenté dans l'annexe B.

7. Inférence

L'inférence bayésienne est une méthode permettant de déduire la probabilité d'un événement à partir de celles d'autres événements déjà évalués. De point de vue mathématique, il s'agit de calculer $P(Y|X)$. C'est la probabilité a posteriori de Y sachant X, tel que X un ensemble d'observations et Y un ensemble de variables décrivent le problème.

En effet, le calcul d'inférence probabiliste étant en général NP-difficile c'est pourquoi il était convenable de le voir de façon complète pour des instances réalisables et incomplète dans les autres cas. D'après [37] [38] les algorithmes d'inférence peuvent être classés en deux catégories : L'inférence exacte, et l'inférence approximative. Le choix d'une catégorie ou d'une autre sera en fonction de la structure du graphe, des distributions de probabilité utilisées et du nombre de données d'apprentissage.

7.1. Algorithmes d'inférence exacte

Les algorithmes d'inférence exacte les plus réputés sont : L'algorithme de passage de messages de Pearl, l'algorithme d'élimination de Bucket et l'algorithme d'arbre de jonction. Ces algorithmes seront détaillés dans la section ci-dessous.

L'algorithme d'élimination des variables

L'algorithme d'élimination des variables ou l'algorithme d'élimination de Bucket en anglais (Bucket Elimination algorithm) décrite dans [40]. Il 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 vaudra 1, ce qui conduit à l'élimination de la variable X_i . On procédera alors à la marginalisation sur des variables restantes ainsi de suite jusqu'à ce que la

distribution soit marginalisée. Le problème de cet algorithme est que l'ordre dans lequel ces variables sont éliminées détermine la quantité de calcul nécessaire pour marginaliser la distribution de probabilités jointes, donc la complexité de l'algorithme.

L'algorithme de passage de messages (Algorithme Pearl)

L'algorithme de passage de messages est plus connu sous le nom de polytree algorithm [41]. C'est la première méthode d'inférence exacte introduite par Kim et Pearl (1983) et Pearl (1985) [42]. Leur principe est d'associer à chaque nœud un processeur qui puisse envoyer des messages à ses voisins jusqu'à ce qu'un équilibre soit atteint un nombre fini d'étapes. Cette méthode ne fonctionne de manière exacte que dans le cas des arbres ou poly-arbres.

➤ Exemple de fonctionnement de l'algorithme dans le cas d'un arbre

- Soit $G(V, E)$ un graphe acyclique orienté (un arbre), où V est l'ensemble des nœuds et E celui des arcs.
- Soit $X = \{X_v : v \in V\}$ un ensemble de variables aléatoires où chaque variable X_v correspond à un nœud v du graphe. Pour chaque nœud $v \in V$ on définit π_v l'ensemble de ses parents dans le graphe.
- Soit θ l'ensemble des probabilités conditionnelles $\{\theta_v\} = \{P(x_v | x_{\pi_v})\}$, $v \in V$ est le couple (G, θ) définissant un réseau bayésien.
- Soit $E \in X$ le sous-ensemble des variables observées E .
- Soit $X_i \in X$ une variable quelconque associée au nœud i de $G(V, E)$.
- Soit N_i l'ensemble des nœuds parents observés de i et D_i l'ensemble des nœuds enfants observés. la figure suivante montre un nœud i , l'ensemble de ses nœuds parents N_i et l'ensemble de ses nœuds enfants D_i , avec N_{X_i} l'ensemble des variables associées aux nœuds N_i et D_{X_i} l'ensemble des variables associées aux nœuds de D_i .

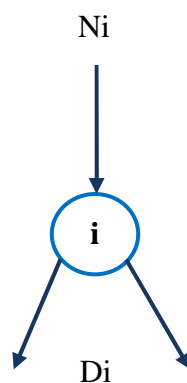


Fig.2. Sous-ensemble d'un arbre : Un nœud, sons parent et ses enfants.

On va alors distinguer deux types de messages λ et π :

$$\lambda(X_i) \propto P(D_{X_i} | X_i) \text{ et } \pi(X_i) \propto P(X_i, N_{X_i})$$

De plus, soit $E_i \in E$ une variable observée, alors :

$$P(X_i | E_i = e_i) \propto \lambda(X_i) \pi(X_i)$$

Où \propto signifie la proportion.

➤ Calcul des messages λ en chaque nœud i

Pour chaque nœud j enfant de nœud i .

I.e. pour chaque variable $X_j \in D_{X_i}$ où $i, j \in V$ et $i \neq j$ on a :

$$\lambda_{X_j}(X_i = x_i) = \sum_{x_j} P(X_j = x_j | X_i = x_i) \lambda(X_j = x_j) \tag{17}$$

$\lambda_{X_j}(X_i = x_i)$ signifie que l'on fait une sommation sur toutes les valeurs possibles x_j de la variable X_j .

Posons $X = \{X_1, X_2, \dots, X_n\}$. Les messages λ se calculent de la façon suivante :

- Si la variable X_i est observée alors $\lambda(X_i)$ est un vecteur de taille égale au domaine de X_i . i.e. le nombre de valeurs possible que la variable X_i peut prendre. Ce vecteur vaut 0 partout sauf à la place de la valeur observée vaut 1.
- Si le nœud i est une feuille de l'arbre, alors le vecteur $\lambda(X_i)$ vaut 1 partout.
- Sinon, $\lambda(X_i = x_i) = \prod_{X_j \in D_{X_i}} \lambda_{X_j}(X_i = x_i)$.

➤ Calcul des messages π en chaque nœud i

- Soit j l'unique nœud parent de i

i.e. Soit $X_j \in N_{X_i}$ où $i, j \in V$ et $i \neq j$.

$$\text{On a : } \pi_{X_i}(X_j = x_j) = \pi(X_j = x_j) \prod_{X_k \in D_{X_i} \setminus X_j} \lambda_{X_k}(X_j = x_j) \tag{18}$$

Posons $X = \{X_1, X_2, \dots, X_n\}$. Les messages π se calculent de la façon suivante :

- Si la variable X_i est observée, alors $\lambda(X_i) = \pi(x_i)$ est un vecteur de taille égale au domaine de X_i . i.e. le nombre de valeurs possible que la variable X_i peut prendre. Ce vecteur vaut 0 partout sauf à la place de la valeur observée vaut 1.

- Si le nœud i est la racine de l'arbre (i.e : si i n'a pas de parent) alors $\pi(X_i) = P(X_i)$
- Sinon, $\pi(X_i = x_i) = \sum_{x_j} P(X_i = x_i | X_j = x_j) \pi_{x_i}(X_j = x_j)$

Par ailleurs, cette méthode a été étendue aux réseaux quelconques pour donner l'algorithme de l'arbre de jonction qui permet de faire l'inférence sur n'importe quel graphe.

L'algorithme de l'arbre de jonction (Clique Tree propagation) [43]

L'algorithme d'arbre de jonction est une généralisation de l'algorithme de passages de Pearl, il permet de faire l'inférence sur n'importe quel type du graphe. On l'appelle aussi algorithme JLO, du nom de ses auteurs: F.V. Jensen, S.L. Lauritzen et K.G. Olesen [44]. Cet algorithme se comporte de la façon suivante:

➤ la phase de construction

La phase de construction nécessite un ensemble de sous-étapes permettant de transformer le graphe initial en un arbre à jonction dont les nœuds sont des regroupements de nœuds du graphe initial. Cette transformation est nécessaire d'une part, pour éliminer les boucles du graphe et d'autre part, pour obtenir un graphe plus efficace quant au temps de calcul nécessaire à l'inférence. Mais elle reste équivalente au niveau de la distribution de probabilité représentée. Elle se fait en trois étapes [45]:

- la moralisation du graphe.
- la triangulation du graphe et l'extraction des cliques qui formeront les nœuds du futur arbre. À rappeler qu'une clique est un sous graphe du graphe G dont tous les nœuds sont connectés deux à deux.
- la création d'un arbre, appelé arbre à jonction.

Cette phase est essentielle, car la qualité de l'arbre de jonction engendré influe directement sur la complexité des calculs lors de la phase de propagation [46].

➤ la phase de propagation

La phase de propagation s'agit de la phase de calcul probabiliste, où les nouvelles informations concernant une ou plusieurs variables sont propagées à l'ensemble du réseau, de manière à mettre à jour l'ensemble des distributions de probabilités du réseau.

- La triangulation du graphe moral
- La construction de l'arbre de jonction

Mais, ces algorithmes souffrent d'un grand problème de la complexité. Tel que, si on a des graphes qui sont fortement connectés, la complexité sera de type exponentiel. Pour pallier ce problème, on peut utiliser des méthodes d'inférence approximatives qui donnent des résultats approchés mais dans un temps fini.

7.2. Algorithmes d'inférence approximative

Lorsqu'il n'est pas possible de faire une inférence exacte, par exemple le cas où les facteurs qui interviennent dans l'algorithme de variables d'élimination ou celui de l'arbre à jonction sont très larges à manipuler efficacement, nous devons chercher une méthode approximative qui permette d'avoir un résultat approximatif mais très proche de celui traité par une inférence exacte.

Par ailleurs, pour certains réseaux bayésiens contient des nœuds discrets et des nœuds continus, il n'existe pas d'algorithme d'inférence exact. Le seul recours pour évaluer des probabilités est alors d'utiliser des algorithmes d'inférence approchés. Trois méthodes différentes ont été proposées : La méthode de propagation cyclique MCMC (pour Markov Chains Monte Carlo) [43] permettant d'échantillonner des variables aléatoires en construisant une chaîne de Markov. L'échantillonnage de Gibbs [47] est une technique qui génère indirectement des variables aléatoires à partir d'une distribution sans nécessiter le calcul de la densité et les méthodes variationnelles [48].

Ces méthodes statistiques approchées ne calculent pas exactement les lois marginales, mais donnent une estimation. Elles permettent donc de traiter des applications de grande taille en temps raisonnable ce qui n'est pas le cas des méthodes exactes.

Pour conclure, on peut dire que l'inférence dans les RB est un problème maîtrisé et suivant les en faisant appelle à des méthodes exactes ou approchées, selon que l'on souhaite privilégier la performance (temps de calcul) ou la précision des résultats.

8. Réseaux Bayésiens adaptés à la classification

Les classifieurs Bayésiens utilisent des méthodes basées sur le théorème de Bayes afin de déterminer les probabilités et d'associer certaines classes à certains objets selon les données d'apprentissage. Le problème de classification peut se traduire par la

minimisation du taux d'erreurs, ce qui peut être formulé mathématiquement en utilisant la règle de Bayes. Dans le cadre de l'apprentissage bayésien nous retrouvons plusieurs types de classificateurs : classificateur optimal de Bayes, classificateur Bayésien Naïf, classificateur de Gibbs et les réseaux bayésiens.

Dans cette partie nous allons présenter le classificateur Bayésien Naïf qui sera utilisé dans nos contributions.

Le classificateur bayésien naïf

Le classifieur bayésien naïf repose sur l'hypothèse que les solutions recherchées peuvent être trouvées à partir des distributions de probabilités dans les données et dans les hypothèses. Cette méthode permet de déterminer la classification d'un exemple quelconque spécifiée en termes de variables. En supposant que les variables $\{x_1, \dots, x_n\}$ de l'espace d'entrée X sont indépendantes les unes des autres et $y \in Y$ tel que $Y = \{0,1\}$ pour la classification binaire. La règle de classification de Bayes s'écrit :

$$C_{N \text{ Bayes}}(\mathbf{x}) = \operatorname{argmax}_{y \in Y} P(x_1, \dots, x_n | y) \times P(y) \quad (19)$$

- C la variable aléatoire représentant la classe.
- x_1, \dots, x_n sont les variables caractéristiques.

Les réseaux Bayésiens naïf, peuvent être représentés par la structure de la figure ci-dessous :

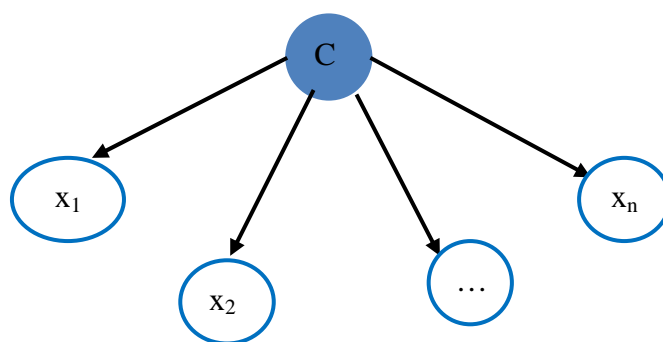


Fig.3. Exemple d'un réseau bayésien naïf

Le principal avantage de ce classifieur est sa vitesse d'apprentissage et ses bonnes performances prédictives. Avec très peu de données, sa précision est bien souvent meilleure que celle d'autres algorithmes. Cette qualité fait que le classifieur Bayésien naïf est assez

souvent utilisé en combinaison avec d'autres algorithmes, comme par exemple les arbres de décision [19].

Pour améliorer la performance du classificateur basé sur l'hypothèse d'indépendance conditionnelle des variables feuilles, Friedman propose d'augmenter la structure du réseau bayésien naïf avec un arc entre les caractéristiques quand c'est nécessaire. Cette structure (**Fig.4**) est appelée réseau bayésien naïf augmenté. Le but de cette amélioration est l'augmentation de pouvoir classifier en recherchant le meilleur arbre.

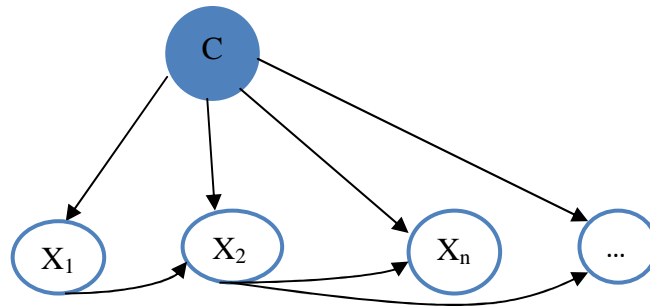


Fig.4. Réseau bayésien naïf augmenté par un arbre

Cependant, il existe une autre approche plus récente appelée les multi-nets [42] qui a apporté deux points importants. Dans le premier point, les relations de causalité et d'indépendances entre les attributs ne sont pas forcément les mêmes selon les différentes modalités du nœud principal « classe ». Le deuxième point, la structure représentant les liaisons entre les attributs pour une classe fixée est beaucoup plus simple que de la structure qui relie toutes les variables.

9. Conclusion

Les réseaux bayésiens font partie de la famille des modèles graphiques. Ils regroupent au sein d'un même formalisme la théorie des graphes et celle des probabilités afin de fournir des outils efficaces autant qu'intuitifs pour représenter une distribution de probabilités jointes sur un ensemble de variables aléatoires. De même leur représentation graphique facilite la compréhension dans un domaine de connaissances.

Dans ce chapitre, nous avons couvert les principaux aspects concernant les réseaux bayésiens. Ainsi, nous avons vu que ces modèles graphiques sont des outils bien adaptés à la description des problèmes de décisions dans l'incertain, en permettant notamment de tenir compte des connaissances du problème à traité. De plus, nous avons également illustré les fondements théoriques des réseaux bayésiens et nous avons 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. À la fin, nous avons cité quelques modèles particuliers des réseaux Bayésiens parmi lesquels : les réseaux bayésiens dynamiques et les classificateurs bayésiens. Le prochain chapitre sera consacré à l'intégration des réseaux bayésiens dans la reconnaissance des visages.

Chapitre III

Détection de la peau du visage

1. Introduction

Nous étudierons dans ce chapitre une technique de reconnaissance de la peau du visage par l'analyse de la couleur et de la texture dans les images. Cette méthode s'appuie sur les réseaux bayésiens naïfs et la matrice de cooccurrence pour définir un modèle de peau capable de différencier les pixels de peau de ceux de non peau. Les expérimentations réalisées sur une base d'images importante montrent l'efficacité et la faibilité de notre approche. Les résultats de cette méthode seront présentés en notant que l'implémentation de cette méthode proposée a été réalisée sous logiciel Matlab et la Toolbox(BNT).

2. Détection de la peau

Détection de la peau [12], dans une image revient à trouver les pixels et les régions de l'image correspondant à la peau. Ce processus est généralement utilisé comme une étape de prétraitement pour trouver des régions qui pourraient correspondre à certaines parties du corps humain. Aussi la présence de peau dans une image est une indication de la présence possible d'humains. La plupart des travaux de recherches dans ce domaine s'appuie sur la couleur [26] [39], et quelques approches y ajoutent également l'utilisation d'informations de texture [49]. La détection des pixels de la peau est une opération relativement simple, peu couteuse en calcul tout en étant efficace. Ces qualités ont encouragé son utilisation en prouvant son efficacité dans des applications telles que la détection des mains, la détection et le suivi du visage et l'analyse des gestes dans divers systèmes d'interaction homme-machine.

La détection de la peau basant sur la couleur est un processus comporte deux phases : la phase d'apprentissage et celle de classification. Concernant la phase d'apprentissage se base théoriquement en trois étapes principales.

- Choisir un espace couleur approprié.
- Collecter une base de données d'images « patches » de peau à partir d'exemples d'images de visage. 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.

- Apprendre les paramètres du classifieur de peau.

Après l'apprentissage, elle vient la phase de classification qui consiste à classifier les pixels de la peau dans la phase d'apprentissage. De même la phase de classification ce fait dans trois étapes principales :

- De convertir l'image dans le même espace couleur que celui qui a été utilisé dans la phase d'apprentissage.
- De classifier chaque pixel en pixel-peau ou pixel-non-peau, en utilisant le classifieur entraîné.

Des fois, on applique 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.

2.1.Approche basée sur la couleur

Dans notre approche, la couleur est l'indice fondamental pour identifier les pixels de peau de ceux de non peau dans une image. Plusieurs espaces de couleur sont utilisés pour étiqueter les pixels comme pixels de couleur de peau : RGB, HSV, YCrCb, ...

Le choix de l'espace de couleur est très important pour la perception des couleurs proches par l'utilisateur. Les images sont souvent représentées et affichées en espace RGB. L'espace convenable pour la classification et celui qui permet de distinguer les composants intensité et la chrominance, puisque l'espace RGB mélange ces derniers donc il n'est pas convenable pour le faire.

Dans notre travail, nous choisirons le modèle « HSV » car, c'est le plus couramment employé dans les logiciels de référence dans le traitement numérique en termes de retouches couleurs. Plusieurs travaux, ont montré que l'espace couleur HSV (Hue, Saturation et Value) permet une bonne détection des pixels de la peau. L'intérêt de cet espace est que ces composantes sont très proches de la perception humaine des couleurs. Il 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 [50] [51] [52]... faisant appel à la détection de la peau grâce a la notion de chrominance et de luminance.

2.2. Approche basée sur la texture

Les méthodes de l'analyse statistique de la texture sont divisées en deux groupes : une analyse de premier ordre comme la moyenne, la variance, ..., et une analyse de second ordre comme la matrice de cooccurrence. Le but de l'analyse de texture est d'extraire les propriétés caractéristiques de l'objet dans une image et de les exprimer sous une forme de vecteur de caractéristiques. La représentation obtenue servira de base aux étapes ultérieures.

Dans le cadre de notre travail, nous nous intéressons aux méthodes de second ordre, plus précisément nous choisissons les matrices de cooccurrence, parce que les statistiques du second ordre prennent en compte la distribution spatiale de deux pixels décrivant ainsi la relation sur le voisinage des pixels.

3. La matrice de cooccurrence

Une matrice de cooccurrence est une matrice de taille $N \times N$, où N est le nombre de niveaux de gris d'une image. Pour un déplacement d (translation), un élément (i, j) de la matrice est défini par le nombre de pixels de l'image de niveau de gris j situés à d d'un pixel de niveau de gris i .

Les matrices de cooccurrence contiennent une masse très importante d'informations, elles sont donc difficilement manipulables, de ce fait, quatorze indices qui correspondent à des caractères descriptifs des textures peuvent être calculés à partir de ces matrices. Nous ne présentons ici que trois de ces indices:

Homogénéité :

$$\sum_i \sum_j \frac{P(i, j)}{1 + (i - j)^2}$$

Energie :

$$\sum_i \sum_j (P(i, j))^2$$

Contraste :

$$\sum_i \sum_j (i - j)^2 P(i, j)$$

4. Classification par les réseaux bayésiens naïfs

Plusieurs méthodes sont proposées pour construire un modèle de couleur de peau, parmi elles les classifieurs bayésiens naïfs, que nous avons déjà définis dans le second chapitre, et que nous utiliserons dans notre application. Ces classifieurs permettent de déterminer les probabilités d'appartenance a posteriori si les densités de probabilité et les probabilités a priori sont connues. De même, il permet d'obtenir le taux d'erreur de classification minimum, qui est l'objectif souhaitable pour tous systèmes de classification.

Les classifieurs bayésiens naïfs s'appuient sur le théorème de Bayes. L'expression mathématique de la formule de Bayes qui prend en considération la probabilité a priori d'apparition des individus des différentes classes et de leurs distributions dans l'espace des descripteurs est donnée comme suit:

$$P(\text{Classe}_k|\bar{X}) = \frac{P(\bar{X}|\text{Classe}_k)P(\text{Classe}_k)}{P(\bar{X})} \quad (1)$$

Avec $P(\text{Classe}_k|\bar{X})$: probabilité a posteriori qu'un pixel de caractéristiques \bar{X} appartient à la classe k.

$P(\text{Classe}_k)$: Densité de probabilité de \bar{X} si la classe est k.

$P(\bar{X})$: Probabilité d'observer \bar{X}

5. Illustration et interprétation des résultats

Notre méthode proposée ici est de construire un modèle paramétrique de distributions de couleur de peau. Nous considérons des vecteurs de forme $c = [H \ S \ V]^T$ qui représentent le composant de chrominance de chaque pixel dans l'espace HSV. Tout d'abord, nous collectons un ensemble des exemplaires de couleur de peau de plusieurs personnes des ethnies différentes telles que les Africains, Européens.... Ensuite, ils sont modélisés par la distribution Gaussienne suivante :

$$\mu_s = \frac{1}{n} \sum_{j=1}^n c_j \quad (2)$$

$$\Sigma_s = \frac{1}{n-1} \sum_{j=1}^n (c_j - \mu_s)(c_j - \mu_s)^T \quad (3)$$

Où c_j est vecteur d'exemplaire extrait

n est nombre de vecteurs

μ_s est le vecteur moyen du modèle

Σ_s est la matrice de covariance du modèle

Grace à ces paramètres, nous pouvons déterminer un pixel qui a une couleur de peau ou de non peau en calculant sa fonction de densité de probabilité dans la distribution Gaussienne.

$$P(c|Peau) = \frac{1}{2\pi|\Sigma_s|^{\frac{1}{2}}} e^{-\frac{1}{2}(c-\mu_s)^T \Sigma_s^{-1}(c-\mu_s)} \quad (4)$$

En utilisant (4), nous avons constuire une image binaire de sortie dans laquelle chaque pixel est classifié en deux classes: classe des pixels de couleur de peau, classe des pixels de couleur non peau. Pratiquement, cette tache est faite de la manière suivante :

Nous construirons au premier lieu, une base de données constituée de 100 photos que nous avons prises pour plusieurs personnes dans les différentes conditions d'éclairages, différentes couleurs de peau et différentes orientations à l'aide d'une caméra. Ensuite, à partir de cette base nous avons extrait des patches (imassettes) de taille 100 * 100 dans le but de construire une base d'apprentissage contenant 100 imassettes peau et 100 imassettes non peau. Nous avons construit aussi une base de tests à partir de plusieurs images contenant des objets d'une grande variabilité (pixels peau et des pixels non peau). Toutes les images d'apprentissage utilisées sont enregistrées en format JPEG. Nous citerons ci-dessous quelques exemples de base de données et les patches extraits à partir de cette base.



(a) exemple extrait de la base de données



(b) exemple des imassettes peau



(c) exemple des imassettes non peau

Fig.1. Exemple des images d'apprentissage

Une fois que nous avons construit cette base, et que nous avons lu et affiché ces images à l'aide du logiciel MATLAB que nous avons utilisé dans notre travail, nous avons passé à la deuxième étape qui consiste à construire un réseau bayésien naïf. La création de ce dernier se fait grâce à la toolbox (BNT) qui contient toutes les commandes construisant les réseaux bayésiens. Tel que, la commande « **mk-bnet** » pour la création de RBN qui prend en entrée le réseau créé et la base d'apprentissage, ensuite il fait l'initialisation des probabilités conditionnelles aléatoirement avec un a priori de Dirichlet.

Les variables de l'espace colorimétrique HSV sont continues, alors, dans ce cas nous avons utilisé la commande « **gaussian-CPD** » afin de faire l'apprentissage de ces variables.

Après la construction du RBN, nous avons passé à une autre étape dite inférence, dans cette étape, nous avons choisi tout d'abord l'algorithme d'inférence à appliquer, dans notre cas nous avons utilisé l'algorithme d'élimination de variables implémenté dans la BNT sous la commande « **var_elim_inf_engine** ». Le but de cette étape est de classifier les attributs par rapport à la classe peau ou non peau. En utilisant la probabilité a posteriori de chaque classe qui s'écrit de la manière suivante :

$$P(C_i/X=x_j) = P(C_i) \prod_{j=1}^p P(x_j / C_i) \quad (5)$$

Avec :

C_i les valeurs de la classe C et X_j représentent les valeurs de vecteur X de paramètres des pixels de l'image, i est l'ordre des classes ; dans notre cas, nous avons deux classes : $i=1,2$ (peau et non peau) et le p c'est le nombre de paramètres utilisés.

Il y a une autre méthode plus simple que l'utilisation de la BNT. Cette méthode consiste à utiliser la commande « **Naïve Bayes .fit** » pour l'apprentissage et la commande « **Predict** » pour l'inférence.

L'étape suivante consiste à tester des images qui appartiennent à notre base mais aussi d'autres qui n'appartiennent pas à cette base.

De même, pour la texture sauf dans ce cas, nous avons converti les images au niveau de gris à l'aide de la commande « **rgb2gray** ». Ensuite nous avons calculé les paramètres de la texture avec la matrice de cooccurrence. Sa commande est implémentée dans Matlab sous « **graycomatrix** », et après avoir calculé les paramètres, nous l'avons récupéré sous forme d'un vecteur « **data** » que nous avons utilisé pour faire l'apprentissage et l'inférence de la même façon que la couleur.

Dans le but de l'amélioration des résultats nous avons fait la fusion des approches (couleur et texture). A la fin nous avons calculé les taux de bonne détection de la matrice de confusion. Les résultats seront montrés ci-dessus.

5.1. Illustration des résultats

- Les réseaux bayésiens naïfs obtenus dans les différentes applications

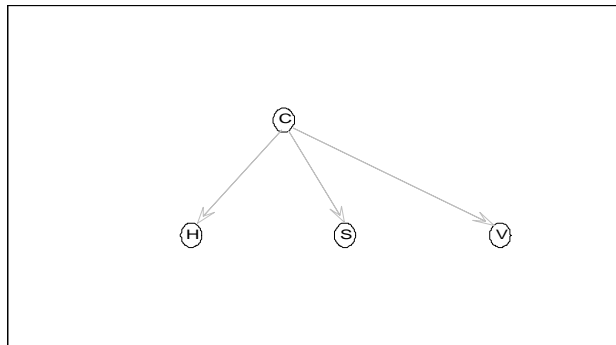


Fig.2. L'architecture de RBN avec les paramètres d'espace couleur HSV

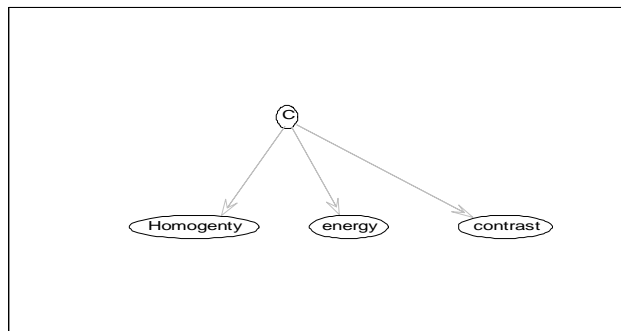


Fig.3. L'architecture de RBN avec les paramètres de texture

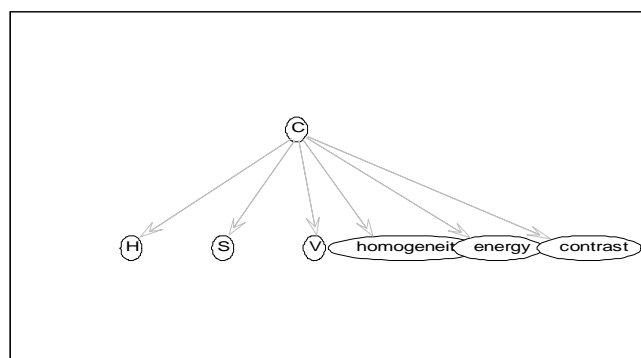
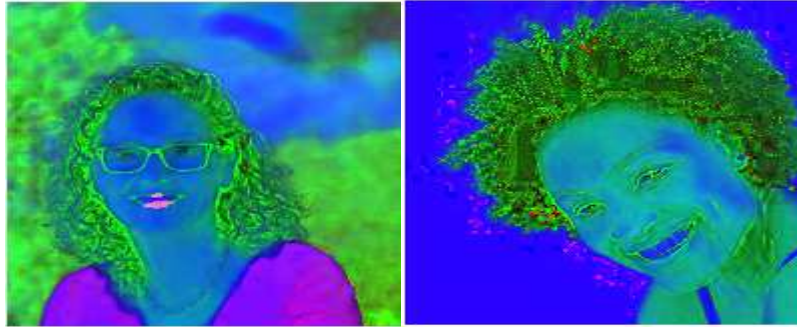


Fig.4. L'architecture de RBN avec les paramètres de la texture et de couleur

Pour les RBN obtenu, le C est le nœud parent représente le nœud classe « peau ou non peau » et les autres nœuds sont les nœuds enfants. par exemple dans le cas de l'espace couleur HSV les nœuds enfants correspondant au trois composantes H, S et V.

➤ *Résultats obtenu*

• **Pour la couleur**



(a)



(b)

Fig .5 . Détection de la peau à partir de sa couleur

(a) Les images HSV, (b) Les images résultantes.

La matrice de confusion obtenue et la suivant :

$$\begin{bmatrix} 68426 & 13182 \\ 7306 & 74302 \end{bmatrix}$$

Le taux de bonne détection est : 87.4473%

Dans le cas, où les images test sont en HSV les résultats obtenus sont montrés ci-dessous, avec le taux de bonne détection égale à 87.45%.

- Pour la texture

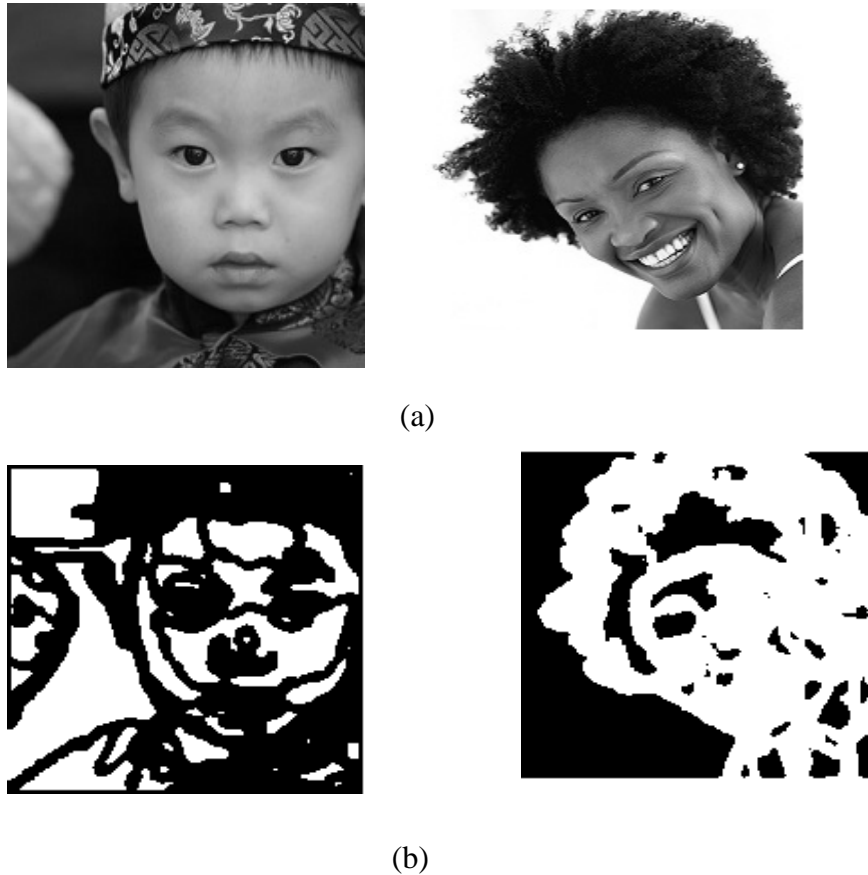


Fig .6 . Détection de la peau à partir des paramètres de la texture.

(a) Les images au niveau de gris, (b) Les images résultantes.

La matrice de confusion obtenue est la suivante :

$$\begin{bmatrix} 3458 & 6743 \\ 1733 & 8468 \end{bmatrix}$$

Le taux de bonne détection, dans ce cas est : 58.4551 %

Les résultats que nous avons obtenus dans cette approche montrent que l'utilisation de la texture pour la détection de la peau n'est pas la meilleure méthode par rapport à la couleur tel. En remarquant la grande différence (58.4551% pour la texture est pour la couleur de taux de bonne détection est 87.4473%).

- Pour la fusion



(a)



(b)

Fig .7 . Détection de la peau en fusion les deux approche

(a) Les images tests originales, (b) Les images résultantes.

La matrice de confusion obtenue est la suivante :

$$\begin{bmatrix} 990928 & 18971 \\ 30579 & 979320 \end{bmatrix}$$

Le taux de bonne détection est : 97 ,5468

A partir de cette application, nous pouvons dire que les RBN donnent des résultats probants, en fusionnant des paramètres de la texture avec celle de la couleur permet d'améliorer les résultats, en remarque que le taux de détection atteint les 97.54%.

À la fin, pour conformer les résultats obtenus nous avons testé les trois approches sur la même image test. Les résultats obtenus sont montré sur la figure suivante :

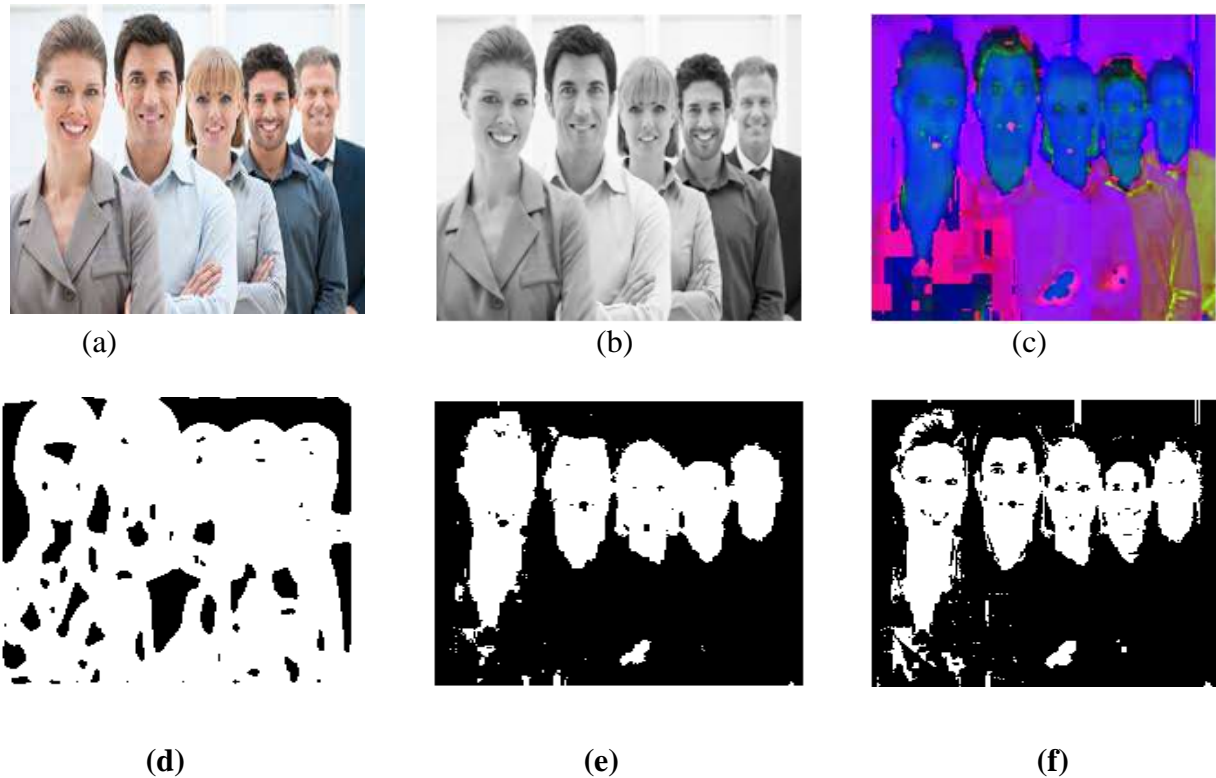


Fig.8. Images test : (a). Image original (b). Image au niveau de gris (c). Image en HSV (d), (e), (f) Les images résultantes par ordre (texture, couleur, la fusion (texture+couleur)).

Pour la texture : le taux de bonne classification =

Pour la couleur : le taux de bonne classification = 87.4473%.

Pour la fusion : le taux de bonne classification =

Nous remarquons que le taux de bonne détections est meilleure lors de la fusion des paramètres couleurs et ceux de texture.

5.2. Interprétation des résultats

Nous avons appliqué les réseaux bayésiens naïfs pour la détection de la peau du visage, pour cela nous avons utilisé les paramètres de la couleur et texture. Et nous avons obtenu le meilleur taux de détection lors d'assembler des deux paramètres (texture et couleur), les petites perturbations sont dues aux conditions d'éclairage et certains objets qui ont des couleurs semblables a celles de la peau humaine, dont laquelle ne peut pas faire la différence même avec l'œil. Nous avons constaté que les paramètres de la texture seuls

donnent des mauvais résultats. Par contre, la couleur est un moyen très efficace pour la détection de la peau. Et en fusionnant la couleur et la texture le résultat est meilleur.

6. Conclusion

L'application que nous avons présentée dans ce manuscrit est une étape préliminaire et d'une grande importance dans tous les systèmes de reconnaissance de visages, car avant de lancer le système pour reconnaître les personnes, une vérification de la présence ou pas d'un visage dans la scène peut être très évidente. Cette étape permettra au système de détourner l'opération de reconnaissance si la scène ne contient aucun visage.

Notre application est basée sur la recherche de la présence de la peau dans l'image, à l'aide d'un classifieur bayésien naïf en se basant sur l'information couleur et texture.

Les résultats obtenus sont satisfaisants, car le taux de bonne détection est important en le comparant à celui obtenu en utilisant la couleur et texture séparément.

Nous avons aussi montré que l'information texture n'est pas suffisante seule mais permet d'améliorer sensiblement les performances d'un détecteur de visages.

Conclusion générale

Nous nous sommes intéressés dans ce mémoire à la reconnaissance d'objet coloré, en appliquons les méthodes graphique probabiliste plus précisément les réseaux bayésiens. Tel que ces derniers constituent une technique d'acquisition, de représentation et de manipulation de connaissances, qu'ils s'agissent de formaliser la connaissance d'expert, d'extraire la connaissance contenue dans des bases de données...ils permettant de représentation des connaissances incertaines et de raisonnement à partir d'informations incomplètes sur des phénomènes complexes. Ces avantages les rends plus utilisé dans plusieurs domaines tels que l'informatique, traitement d'images, la santé, la défense, la sécurité, etc.

L'objectif de notre travail et d'utiliser les réseaux bayésiens dans le domaine de sécurité plus précisément la reconnaissance faciale pour identifier le visage à partir de la peau on s'appui sur les deux techniques d'identification couleur et texture. La raison de ce choix est que cette approche s'affranchit de la plupart des difficultés liées à la détection de visage précédemment citées (posture, expression faciale, orientation,...etc.). Pour cela, nous avons utilisé l'espace couleur HSV pour ses capacités de distinguer les composants intensité et chrominance et on a aussi extraits à partir des matrices de cooccurrence trois attributs de texture: contraste, énergie, et homogénéité.

Pour l'extraction des descripteurs pertinents, nous avons proposé et dressé un ensemble d'algorithmes de sélection des attributs qui sont évalués en précision et une mesure de score proposée pour la recherche d'images et aussi par les taux de classification et matrices de confusion dans le cas de la classification supervisée. Tel que les algorithmes d'inférences et d'apprentissages des réseaux bayésiennes.

Les tests effectués sur quelques images, sont montré des bons résultats dans la quelle nous avons utilisé l'approche couleur, par contre pour la texture il donne des résultats mais ils ne sont pas vraiment suffisant. Pour remédier à ce problème on a pensé de fusionné le deux approche. A cette fin on a trouvé des résultats super en comparaison avec les deux approches utilisé séparément. Tel que, le taux de bons détection est améliorée de 10.0995% par rapport a celui de la couleur et nous ne parlerons pas pour celui de la texture.

De plus, nous avons opté pour l'utilisation des classifieurs bayésienne naïf comme des bons classifieurs pour la détection de la peau en basant sur des caractéristiques couleurs et texture afin de trouver les visages présents dans l'image qui étayent l'hypothèse de la présence de personnes. Par ailleurs, cette approche présente une relative simplicité de mise en œuvre et exhibe une vitesse d'exécution compatible avec le contexte applicatif de nos travaux.

Références bibliographiques

- [1] : **M. DEMRI** « Multimodal biometric fusion using evolutionary techniques » mémoire de magister en informatique, Algérie, 2012.
- [2] : **S. GUERFI ABABSA** « Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D », thèse de doctorat a l'université Evry Val d'Essonne, 2008.
- [3] : **A. CHAARI** « Nouvelle approche d'identification dans les bases de données biométriques basée sur une classification non supervisée », Thèse de doctorat en informatique, à UEVE et université de la Manouba, 2009, IBISC.
- [4] : **F. PERRONNIN & J-L.DUGELAY** « Introduction à la biométrie Authentification des individus par traitement audio-vidéo » Traitement du signal 2002-volume 19n^o4, Institut Eurecom, multimédia communication département 2229, route des crêtes, B.P.193, 06904 Sophia Antipolis cedex perronni. <http://www.eurecom.fr/.image>.
- [5] : **Groupe de travail sur la sécurité de l'information et la vie privée** « Technologies fondées sur la biométrie » Editions OCDE ,2005. <http://dx.doi.org/10.1787/232025043655>.
- [6] : **P. JOYEUX** « La biométrie, ses différentes techniques » Conservatoire National des arts et Métiers, centre de Lyon-hone-Alpes, 2006.
- [7] : **Commission techniques de sécurité physique** « Techniques de contrôle d'accès par biométrie », 2003 <http://www.clusif.esso.fr>
- [8] : **I. DEHACHE & L.S. MESLATI** « Une approche multimodale pour la vérification biométrique », Courrier du savoir- N^o12, Octobre2011 PP.79.86, Université Mohamed Khider, Biskra, Algérie, 2009.
- [9] : **B.E. BOSER & I.M. GUYON & V.N. VAPNIK** « A training algorithm for optimal margin classifiers » In fifth annual workshop on computational learning Theory, pages 144-152, Pittsburgh, ACM.1992
- [10] : **H. SHERIF** « L'identification biométrique dans le commerce électrique », Editions Eyrolles©, 2002.

[11] : **Commissariat à la protection de la vie privée du canada** « Reconnaissance faciale automatisée dans les secteurs public et privée », mars 2013, www.priv.gc.ca.

[12] : **F. SOYEZ** « Reconnaissance faciale : plus vraiment de la science fiction. CENT France. Fiction-39786494.htm> Janvier 2013.

[13]: **R. YAPP** « Brazilian police to use robocop-style glasses at world cup » The telegraph <http://www.Telegraph.co.uk/news/worldnews/southamerica/brazil/8446088>>, avril 2011.

[14] : **F. ABDAT** « Reconnaissance automatique des émotions par données multimodales : expressions faciales et signaux », Thèse de doctorat à l'université de METZ en automatique, 2010.

[15] : **M.V. WAMBEKE** « Reconnaissance et suivi de visage et implémentation en robotique temps réel », Mémoire de master en génie biomédical, Ecole polytechnique de LOUVAIN, 2009-2010.

[16] : **S. BARRAT** « Modèles graphiques probabilistes pour la reconnaissance de formes », Thèse de doctorat à l'université Nancy 2 en informatique, 2009.

[17] : **V. CLAVEAU & E.KIJAK & O.FERRET** « Explorer le graphe de voisinage pour améliorer les thésaurus distributionnels », 21^{ème} traitement automatique des langues naturelles, Marseille, 2014.

[18] : **E.M. DUPAS** « Algorithme des K plus proches voisins pondérés et application en diagnostic », 42^{ème} Journée de statistique, 2010 Marseille, France, <inria-00494814> <https://hal.inria.fr/inria-00494814>.

[19]: **B.E.BOSER& I.M.GUYON& V.N.VAPNIK** « A training algorithm for optimal margin classifiers », in fifth annual workshop on computational learning theory, pages 144-152, Pittsburgh, ACM. 1992.

[20]:**M.DAHMANE** “Analyse de mouvements faciaux à partir d’images vidéo”, Thèse de doctorat en informatique à l’université de Montréal, 2011.

[21] : **A. DJEFFAL** « utilisation des méthodes support vector machine (SVM) dans l’analyse des bases de données». Thèse de doctorat en informatique 2012 université Mohamed khider, Biskra.

[22] : **Y.JIE & WAIBEL.A.SARASOTA** « A real time face tracker », Third IEEE work shop on applications of computer Vision, (WACV 96), P142-147, 1996.

[23] : **M.CHOPIN & P.H.WUILLEMIN** « Optimisation de l'inférence dans les réseaux bayésiens Dynamiques », LIP6, université de paris, France. ROADEF, 2010, Toulouse.

[24] : **M.V.WAMBEKE**« Reconnaissance et suivi de visages et implémentation en robotique temps-réel », mémoire de master ingénieur civil en génie biomédical, 2010.Ecole polytechnique de Louvain.

[25] : **T.P. NGUYEN** « détection de visage dans images de couleur en utilisant une caractéristique invariante » 3rd International conference : sciences of electronic March 27-31, 2005- Tunisia.

[26] : **P. JULIAN & V.CHARVILLAT & C.DEHAIS & F.LAUZE** « de l'intérêt de la texture pour la segmentation du visage », la pyrénéenne, labège France. Dans compression et représentation des signaux audiovisuels (coresa), 2009.

[27]: **Z. HAMMAL & I. COUVREUR & A.CAPLIER & M. ROMBANT** « Facial expression classification: An approach based on the fusion of facial deformations using the transferable belief model. Int. J.Approx. Reason, 46(3): pp.542-567.

[28] : **O. PARENT & J. EUSTACHE** « Les réseaux bayésiens a la recherche de la vérité », Cours cognition et connaissance de master II , Université CLAUDE BERNARD LYON 1, 2006/2007.

[29] : **M. EMBARKI** « Etude comparative entre réseaux bayésiens et autre méthode de classification appliqué sur une base cardiologue » mémoire de master en informatique, 2012.

[30] : **O. PARENT & J. EUSTACHE** « Les réseaux bayésiens : A la recherche de la vérité », cours de master II recherche connaissance et raisonnement, 2007, Université Claude Bernard Lyon 1.

[32] : **L.MAMERI & C.LARABI** « Réseaux bayésiens en traitement d'images Application à la détection de la peau » mémoire de master académique en automatique, 2012, Alger.

[33]: **Y. AMANULLUH &PH. LERAY** « IMMPC; une approche locale pour l'apprentissage incrémental de la structure des réseaux bayésiens", Equipe connaissance et décision, laboratoire d'informatique de Nantes Atlantique UMR6241.

[34]: **O.FRANCOIS & PH.LERAY** « Evaluation of structure learning algorithms for Bayesian networks » INSA Rouen. Laboratoire PST-FRECNRS 2645 BP 08-AV de l'université, 76801 ST-Etienne du Rouvray cedex.

[35]: **Ph. LERAY** « réseaux bayésiens : Apprentissage et modélisation de systèmes complexes » maitre de conférences LITIS-EA U015, 2006.

[36]: **C. Aaron** : « Algorithme EM et classification non supervisée », Thèse de doctorat d'Université Paris I. 2001.

[37]: **P. NAIM & P. H.WUILEMIN & PH. LERAY & O. POURRET & A. BECKER** « Réseaux bayésiens » 3ème édition, EYROLLES, 1999, 2004, 2008.

[38]: **N. JAOUA OTHER** « Estimation bayésienne non paramétrique de systèmes dynamiques en présence de bruits Alplo-stables », Ecole centrale de LILLE, 2013, French « NNT : 2013 ECLI0003 » tel-00929691.

[39]: **B. Bruno** « réseaux bayésiens » 2014.

[40]: **R. DECHTEL** « Bucket elimination: a unifying framework for structur-driven inference ».Irvine, California USA.92717.

[41]: **F. OLIVIER** « De l'identification de structure de réseaux bayésiens à la reconnaissance de formes à partir d'informations complètes ou incomplètes », Thèse de doctorat en génie traitement du signal, informatique, 2006.

[42]: **J.H.KIM & J.PEARL** « A computational model for combined causal and diagnostic reasoning in inference system. In proceedings IJCAI-83? Pages 190-193, Karlsruhe, Germany, 1983.

[43]: **M. CHOPIN & P.H WUILLEMIN** « optimisation de l'inférence dans les réseaux bayésienne dynamiques » LIP6 ; Université de paris 6 ; 104 ; avenue de président Kennedy, 75016 Paris, France.

[44]: **A. ZAIDI** « Intégration des Réseaux Bayésiens et bond graphs pour la supervision des systèmes Dynamiques », Thèse de doctorat en génie électrique à l'université LILLE 1, 2012.

[45]: **S. BARRAT** « Modèles graphiques probabilistes pour la reconnaissance de formes », Thèse de doctorat en informatique, Université Nancy 2, 2009.

[46] : M. CHOPIN & P.H. WUILLEMIN « Optimisation de l'inférence dans les réseaux bayésiens dynamiques » LIP6, université de paris 6, 104, avenue du président Kennedy, 75016 paris, France .Roadff, 2010, Toulouse.

[47] : V. VINCENT « Echantillonnage de gibbs avec augmentation de données et imputation multiple ».mémoire pour l'obtention du grade de maître es arts (M.A.), Université Laval, 2006.

[48] : Ph. LERAY « Réseaux bayésiens pour la modélisation et la découverte de connaissances » cours du master, 2012.

[49] : N.T. Tankam & A. Dipanda & E. Tonyé & A. Akono « Méthode optimisée de classification d'images satellitaires radar RSO par usage des paramètres de texture d'ordre supérieur. Application à la mangrove littorale Camerounaise», ENSP, Yaoundé, Cameroun et Laboratoire Electronique et Informatique et Image (LE2I), Université de Bourgogne, Dijon, France.

[50]: D. SAXE & R. FOULDS "Toward robust skin identification in video images", 2nd International Face and Gesture Recognition Conference, 1996.

[51]:R. Kjeldsen & J. kinder "Finding skin in color images", Proc. 2nd Int. Conf. on Automatic Face and Gesture Recognition, pp. 312-318, Vermont, 1996.

[52]: M. Fedias « Combinaisons de données d'espaces couleurs et de méthodes de vérification d'identité pour l'authentification de visages », Thèse de doctorat en automatique, Biskra.

[53] : M. HAMMAMI « Modèle de peau et application à la classification d'images et au filtrage des sites Web », Thèse de doctorat en informatique, Ecole national de Lyon, 2005.

[54]: M. IGGANE, D. MAMMASS, A. KHENCHAF, J.CH. CEXUS « Segmentation d'images sonar par matrice de cooccurrence », 5th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, 2009, TUNISIA.

[55]: M. Jacques-A. Landry « Méthodes d'extraction de caractéristiques et de segmentation non supervisée pour textures naturelles », 2009, Montréal.

Annexe A

Les généralités

A.1. Variance - écart-type

La variance et l'écart-type d'une variable aléatoire sont définis respectivement par :

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2 \quad \text{et} \quad \sigma(X) = \sqrt{\text{Var}(X)}$$

Pour une variable aléatoire X et tout réel λ on a :

$$\sigma(\lambda X) = \lambda^2 \text{Var}(X) \quad \text{et} \quad \sigma(\lambda X) = |\lambda| \sigma(X)$$

A.2. Moralisation

La moralisation d'un GAD $G = (V, E)$ est obtenu en désorientant les arcs de ce graphe et en mariant les parents d'un même nœud. Un graphe moral noté $G^m = (V, E^m)$ a la propriété de créer une clique pour chaque nœud et ses parents.

A.3. Triangulation

Un graphe non dirigé est triangulé si et seulement si tout cycle de longueur supérieure à 3 possède une corde (c'est à dire une arête reliant deux nœuds non adjacents dans le cycle). Le graphe de jonction minimal d'un graphe triangulé G^t est un arbre à jonction.

A.4. Représentation graphique de la causalité

Un réseau bayésien ou réseau probabiliste a pour objectif d'acquérir, de représenter et d'utiliser la connaissance. Il est constitué de deux composantes :

- Un graphe causal, orienté, acyclique, dont les nœuds sont des variables d'intérêt du domaine, les arcs 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 distributions locales de probabilités qui sont les paramètres du réseau. Pour chaque nœud on dispose d'une table de probabilités $P(\text{variable} \mid \text{parents})$ qui représente la distribution locale de probabilité. Il faut remarquer que l'état de

chaque nœud ne dépend que de l'état de ses parents. Il s'agit de la représentation quantitative de la connaissance.

A.5. Représentation probabiliste de la causalité

La théorie des probabilités a pour objectif de modéliser des expériences pour lesquelles plusieurs résultats peuvent se réaliser mais leur réalisation ne peut pas être déterminée à l'avance.

Avec la représentation graphique de la causalité on peut connaître la direction de circulation de connaissances dans le graphe mais on ne peut pas connaître la quantité de circulation de connaissances. Alors, il faut une représentation probabiliste associé à ce graphe. Avec une relation causale : $A \Rightarrow B$ on peut représenter la quantité de cette relation par la probabilité conditionnelle : $p(B|A)$ tout en respectant les règles de transposition des graphes en termes de probabilités.

Annexe B

Quelques rappels sur les probabilités

B.1. Définition principale de la probabilité

La probabilité $P(A)$ d'un évènement A est un nombre dans l'intervalle $[0, 1]$ tel que :
 $P(A) = 1$ si l'évènement A est certain.

$P(A) = 0$ si cet évènement est certain et qu'il ne soit pas produit.

Si A et B sont des évènements mutuellement exclusifs et couvrent tous les évènements possibles, on a :

$$P(A) + P(B) = 1$$

B.2. Probabilité conditionnelle

Soit A et B deux évènements. Si $P(A) > 0$, on pose :

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

L'expression $p(B|A)$ se dit « probabilité conditionnelle de B sachant que A est réalisé »

B.3. Probabilité jointe

Soient X et Y deux variables aléatoires continues. Alors leurs fonction de distribution jointe est définie par :

$$F(x, y) = P(X \leq x, Y \leq y)$$

Dans le cas générale, la fonction de distribution jointe de n variables X_1, X_2, \dots, X_n est définie par :

$$F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

B.5. Arbre à jonction

Un arbre à jonction dans un réseau bayésien avec des variables aléatoires X est un graphe acyclique non dirigé $T = (V, E)$ avec V l'ensemble des nœuds et E l'ensemble des arcs. Chaque nœud $C_i \in V$ dans le graphe est appelé clique, il est associé à des variables $X_i \subseteq X$. On nomme $S_{ij} = C_i \cap C_j$ le séparateur des cliques C_i et C_j . Chaque clique C_i est associée à un facteur $\Phi(C_i)$ appelé potentiel de clique et chaque séparateur S_{ij} est associé à un facteur $\Phi(S_{ij})$ appelé potentiel de séparateur.

B.4. Indépendance conditionnelle

Soient deux variables aléatoires X_1 et X_2 . On dit que X_1 et X_2 sont indépendants conditionnellement à X_3 , si l'une des propriétés équivalentes suivantes est vérifiée :

$$P(X_1|X_2, X_3) = P(X_1|X_3)$$

$$P(X_1, X_2|X_3) = P(X_1|X_3).P(X_2|X_3)$$

On notera que $P(X_1, X_2) = P(X_1 \text{ et } X_2)$ est la probabilité jointe alors que la probabilité conditionnelle est $P(X_1|X_2) = \frac{P(X_1, X_2)}{P(X_2)}$, définie seulement lorsque $P(X_2) \neq 0$.

B.5. D-séparation (directed separation)

On dira que X et Y sont d-séparés par Z si pour tous les chemins entre X et Y , l'une au moins des deux conditions suivantes est vérifiée :

- Le chemin converge en un nœud V , tel que $V \neq Z$, et V n'est pas une cause directe de Z .
- Le chemin passe par Z , et il est soit divergent, soit en série au nœud Z .

On a prouvé que si X et Y sont d-séparés par Z , alors X et Y sont indépendants conditionnelle à Z . Ce résultat est très important, car il permet de limiter les calculs de probabilités grâce à des propriétés du graphe. Supposons que X et Y soient d-séparés par Z , et que Z soit connu. Supposons, par ailleurs, qu'on vienne de calculer $p(X|Z)$. Si une nouvelle information sur Y est alors connue, ce résultat me permet de conserver le calcul de $p(X|Z)$ comme valeur de $p(X|Z, Y)$.

B.6. Distribution de Dirichlet

La loi de Dirichlet, souvent notée $\text{Dir}(\alpha)$, elle est une famille de lois de probabilités continues. Cette loi (distribution) est paramétrée par le vecteur α de nombres réels positifs.

Tel que, soient $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_r$ tous strictement positifs. Une distribution de Dirichlet d'exposants $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_r$ possède une densité de probabilité de la forme suivante :

$$\text{Dir}(\theta \mid \alpha_1, \alpha_2, \alpha_3, \dots, \alpha_r) = \frac{\Gamma(\sum_{i=1}^r \alpha_i)}{\prod_{i=1}^r \Gamma(\alpha_i)} \prod_{i=1}^r \theta^{\alpha_i - 1}$$

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

Annexe C

Présentation de logiciel Matlab et la BNT

C.1. Présentation de logiciel Matlab

MATLAB (**MA**Trix **LAB**oratory) est un logiciel de calcul numérique, dont les objets privilégiés sont les matrices (réelles, complexes), les vecteurs (réels, complexes) et les scalaires. Les opérations sous MATLAB sont donc naturelles au sens de ces objets. La simplicité d'utilisation de MATLAB en fait un outil pédagogique indispensable à la compréhension de cours, comme le traitement d'image, traitement de signal, l'algèbre linéaire ... Pour toutes les commandes de MATLAB, il existe une documentation directement accessible à partir de la machine : cette aide est réalisée par la commande : *help commande*.

C.2. Bayes Net Toolbox (BNT)

La boîte à outils de Bayes net est une bibliothèque open -source de MATLAB pour les modèles graphiques dirigés. BNT supporte de nombreux types de nœuds (distributions de probabilité), l'inférence exacte et approximative, l'apprentissage des paramètres ou de structure et les modèles statiques ou dynamiques. Dans ce fait, la BNT est largement utilisé dans la recherche et l'enseignement, sa due a leur possibilité de récupéré 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.

Pour charger la BNT dans Matlab nous suivons les instructions suivantes :

Ouvrir le logiciel matlab, aller dans file, Set path, Add with sub folders, la BNT, ok, save, close.

La BNT s'avère comme un outil indispensable et très pratique pour la manipulation des réseaux bayésiens, car elle à la possibilité d'estimer tout les paramètres de ces dernières. La BNT nous mit à notre disposition toutes les fonctions utilisées dans l'inférence et l'apprentissage des paramètres (probabilité a posteriori, l'algorithme d'arbre de jonction,...).

Résumé

La reconnaissance faciale est une technique biométrique la plus adaptée dans les dernières années, non seulement pour son efficacité, mais aussi pour sa possibilité de diminuer le temps de calcul grâce à sa simplicité et la robustesse lorsqu'on travaille avec une grande base de données.

Dans notre manuscrit, nous avons présenté un système de détection de visages à partir des images fixes, notre système est testé, avec une variété d'images contenant plusieurs personnes sans conditions sur l'éclairage et sans conditions sur l'angle de rotation du visage dans l'image, il est basé sur un réseau bayésien (classifieur bayésien naïf). Ce réseau s'appuie sur la méthode de détection de couleur en utilisant l'espace HSV ainsi, la méthode de détection de texture en utilisant la matrice de cooccurrence.

A cet effet, nous avons montré que ce classifieur bayésien donne des résultats probants malgré sa structure simplifiée.

Mots-clés : reconnaissance faciale, détection de la peau, Couleur, texture, le réseau bayésien naïf, la matrice de cooccurrence.

Abstract

Facial recognition is a technique best suited biometrics in recent years, not only for its effectiveness but also for its ability to reduce the time to calculate due to its simplicity and robustness when working with a large database.

In our manuscript we present a face detection system from still images, our system is tested with a variety of pictures containing several people were unconditionally and lighting requirements without the facial rotation angle in the image. It's based Bayesian networks (naive Bayesian classification). The Bayesian networks consisting of the color detection method and the detection method using skin texture the co-occurrence matrix.

To this end, we have shown that this Bayesian classifier gives convincing results despite its simplified structure.

Key words: Facial recognition, Skin detection, Color, Texture, Naive Bayesian Network, co-occurrence matrix.