

وزارة التعليم العالي و البحث العلمي  
**MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA  
RECHERCHE SCIENTIFIQUE**

جامعة فرحات عباس – سطيف  
**UNIVERSITE FERHAT ABBAS - SETIF  
U.F.A.S. (ALGERIE)**

## **MEMOIRE**

Présenté à la Faculté des Sciences de l'Ingénieur  
Département d'électronique

Pour l'obtention du Diplôme de

## **MAGISTERE**

**Option: CONTRÔLE**

Par

**Mr. BARKI Hicham**

THEME

## **DETECTION ET RECONNAISSANCE DE VISAGE.**

Soutenu le:....., devant la commission d'examen:

KHELLAF Abdelhafid  
BENMAHAMMED Khier  
MOHAMMADI Tayeb  
BENAOUDA Abdelhafid  
MOUSSAOUI Abdelouahab

Prof. à l'université de Sétif  
Prof à l'université de Sétif  
Prof. à l'université de Sétif  
Dr. à l'université de Sétif  
M.C. à l'université de Sétif

Président  
Rapporteur  
Examineur  
Examineur  
Examineur

# Remerciements

Je remercie et je loue tout d'abord dieu de m'avoir aidé à défier tous les obstacles, afin de compléter ce modeste travail.

C'est avec un immense plaisir que j'exprime mes remerciements, les plus chaleureux à mon encadreur, le Professeur BENMAHAMMED Khier qui a toujours suivi ce travail avec intérêt et qui l'a témoigné par des encouragements de toutes sortes.

Je remercie très vivement le président et les membres du jury qui ont accepté de juger ce modeste travail.

Enfin, pour tous ceux qui, de près ou de loin, ont rendu ce travail possible, je dis pour la nième fois

Merci.

# Dédicaces

Je dédie ce modeste travail :

A Mes parents, à mes frères et sœurs.

A toute ma famille.

A tous mes enseignants et enseignantes.

A tous mes amis.

<b>Table des matières:</b>	
<b>Introduction générale</b>	<b>1</b>
<b>Chapitre 1: La biométrie</b>	<b>3</b>
1.1. Introduction	4
1.2. Un peu d'histoire	4
1.3. Qu'est ce que la Biométrie ?	5
1.4. Pourquoi utiliser la biométrie ?	6
1.5. Les systèmes biométriques	7
1.5.1. Les empreintes	7
1.5.2. La main	9
1.5.3. La voix	9
1.5.4. La signature	10
1.5.5. Le clavier	11
1.5.6. L'iris	12
1.5.7. La rétine	14
1.5.8. Le visage	15
1.6. Conclusion	16
<b>Chapitre 2: La détection de visage</b>	<b>17</b>
2.1. Introduction	18
2.2. L'évolution de la recherche sur la détection de visage	19
2.3. L'approche basée sur les traits	20
2.3.1. L'analyse de bas niveau	22
2.3.1.1. Les bords	22
2.3.1.2. L'information des niveaux de gris	23
2.3.1.3. La couleur	24
2.4. L'approche basée sur l'image	26
2.4.1. Méthodes des sous espaces linéaires	27
2.4.2. Réseaux de neurones	33
2.4.3. Approches statistiques	36
2.4.4. Evaluation comparative	39
2.5. Conclusion	41
<b>Chapitre 3: La reconnaissance de visage</b>	<b>42</b>
3.1. Introduction	43
3.2. Reconnaissance de visage à partir d'images statiques	47
3.2.1. Les étapes clés avant la reconnaissance du visage: La détection du visage et l'extraction des traits	48
3.2.1.1. Segmentation/Détection	48
3.2.1.2. Extraction des traits	49
3.2.2. Reconnaissance à partir d'images d'intensité	51
3.2.2.1. Approches holistiques	55
3.2.2.1.1. L'analyse en composantes principales	55
3.2.2.1.2. Les autres représentations	64
3.2.2.2. Les approches d'assortiment structurelles basées sur les traits	65
3.2.2.3. Les approches hybrides	68
3.3. Conclusion	73
<b>Chapitre 4: Travail et contribution</b>	<b>74</b>
4.1. Introduction:	75
4.2. Détection de visage	75
4.2.1. Première implémentation	75
4.2.1.1. Première étape: la détection des yeux et des candidats de visage	75

4.2.1.1.1. La recherche des segments analogues aux yeux	76
4.2.1.1.2. La détermination des candidats de visage	80
4.2.1.2. Deuxième étape: la vérification des candidats de visage par projection dans l'espace de visage de la PCA	80
4.2.1.2.1. La création de l'espace de visage	80
4.2.2. Deuxième implémentation	82
4.2.2.1. Première étape: développement du schéma de segmentation et des contraintes géométriques	83
4.2.2.1.1. Développement du schéma de segmentation	83
4.2.2.1.2. Amélioration de la segmentation par des contraintes géométriques	91
4.2.2.2. Deuxième étape: La vérification de la présence du visage dans les régions de peau segmentés	92
4.3. Reconnaissance de visage	93
4.3.1. L'analyse en composantes principales (Principal component Analysis, PCA)	93
4.3.1.1. L'algorithme PCA	95
4.3.1.1.1. Processus de reconnaissance	97
4.3.2. La méthode des visages propres des images réduites (Reduced Image Eigenfaces, RIE)	98
4.4. Expérimentations et résultats	100
4.4.1. Détection de visage	100
4.4.1.1. Première implémentation	100
4.4.1.2. Deuxième implémentation	101
4.4.2. Reconnaissance de visage	102
4.4.2.1. L'analyse en composantes principales (PCA)	102
4.4.2.1.1. Résultats obtenus avec la base de données BioID	102
4.4.2.1.2. Résultats obtenus avec la base de données ORL	104
4.4.2.1.3. Comparaison de distances	104
4.4.2.2. Résultats obtenus avec la méthode RIE	105
4.5. Conclusion	105
<b>Annexe A: Mesures de dissimilitude</b>	<b>107</b>
A.1. Mesures de dissimilitude	108
A.1.1. Variables numériques	108
A.1.1.1. Distance Euclidienne	109
A.1.1.2. Distance de Manhattan (City-block)	109
A.1.1.3. Distance de Chebyshev	110
A.1.1.4. Distance de Minkowski	111
A.1.1.5. Distance quadratique	111
A.1.1.6. Distance de Canberra	111
A.1.1.7. Distance non linéaire	111
A.1.1.8. Séparation angulaire	111
A.1.2. Variables nominales et ordinales	112
A.1.3. Variables binaires	112
A.1.3.1. Simple matching coefficient	113
A.1.3.2. Russell et Rao	113
A.1.3.3. Jaccard	114
A.1.3.4. Czekanowski	114
A.2. Sommaire	114
<b>Annexe B: Rappels d'Algèbre linéaire</b>	<b>115</b>

<b>B.1. Définitions et propriétés fondamentales</b>	<b>116</b>
<b>B.2. Sommaire</b>	<b>119</b>
<b>Conclusion et perspectives</b>	<b>120</b>
<b>Références</b>	<b>122</b>

# **Introduction générale**

La détection de visage dans les images est une tâche principale dans plusieurs applications basées sur la vision, telles que les systèmes biométriques, la conférence vidéo et les interfaces homme machine. Cependant, plusieurs problèmes difficiles causés par la diversité des variations, telles que les races humaines, l'illumination, les expressions faciales, les tailles des visages, les position de la tête (rotations hors plan), les occlusions, les inclinaisons de la tête (rotations dans le plan) et ainsi de suite, empêchent les chercheurs de développer des systèmes efficaces de détection de visage. L'objectif de la détection de visage est de déterminer la présence ou non de visages dans une image et de donner, s'ils sont présents, les positions de ces visages et leur étendu. La détection de visage est un important module de prétraitement pour une reconnaissance de visage efficace.

La reconnaissance de visage est l'un des secteurs de recherche les plus actives dans la vision par ordinateur et la reconnaissance de formes, avec des applications pratiques qui incluent l'identification, le contrôle d'accès et les interfaces homme machine. La tâche de la reconnaissance de visage consiste à comparer une image d'entrée avec une base de données contenant un ensemble d'images de visage avec des identités connues et à identifier le sujet auquel appartient le visage d'entrée. Cependant, une simple implémentation est difficile parce que les visages font preuve de variations significatives dans leur apparence, ces variations sont dues aux conditions d'acquisition des images, à l'illumination, à la position des visages et à l'âge.

Ce mémoire traite les deux problèmes de la détection et de la reconnaissance de visage, il est organisé comme suit:

**Chapitre 1:** Ce chapitre donne un aperçu des technologies biométriques utilisées actuellement ou en cours de développement, telles que les empreintes, le visage, la rétine, ..., et des avantages qu'ils peuvent apporté par rapport aux techniques classiques pour l'identification telles que les mots de passe.

**Chapitre2:** Ce chapitre fait le tour de la détection de visage, en exposant les approches utilisées pour accomplir cette tâche, à savoir l'approche basée sur les traits et l'approche basée sur l'image.

**Chapitre 3:** Ce chapitre s'intéresse à la reconnaissance de visage, à ses applications potentielles, mais se limite aux approches travaillant avec des images statiques, telles que les approches holistiques, les approches d'assortiment (ou structurelles) basées sur les traits et enfin les approches hybrides.

**Chapitre 4:** Ce chapitre fournit les détails techniques et les développements nécessaires aux travaux réalisés (implémentations ou réalisations). Pour la détection de visage, on a proposé deux méthodes, la première est basée sur l'extraction des yeux et la vérification des candidats de visage par la suite, la deuxième méthode est basée sur un modèle de segmentation couleur des pixels de la peau suivie par une vérification des candidats de visage par recherche des yeux. Pour la reconnaissance, on a implémenté la technique des visages propres (PCA) sur deux bases de données différentes (ORL et BioID) avec plusieurs configurations, on a ensuite comparé des mesures de distances différentes et on a implémenté la méthode des visages propres des images réduites (RIE). Enfin, on a présenté les résultats obtenus et les conclusions tirées.

**Annexe A:** Cette annexe fournit les détails des mesures de dissimilitude utilisées tout au long de ce mémoire.

**Annexe B:** Cette annexe fournit un rappel des notions élémentaires de l'Algèbre linéaire.

**Conclusion et perspectives:** Ce mémoire s'achève par des conclusions et les perspectives dans le futur.

# Chapitre 1: La biométrie



### **1.1. Introduction:**

La croissance internationale des communications, tant en volume qu'en diversité (déplacement physique, transaction financière, accès aux services...), implique le besoin de s'assurer de l'identité des individus. L'importance des enjeux, motive les fraudeurs à mettre en échec les systèmes de sécurité existants.

Il y a donc un intérêt grandissant pour les systèmes d'identification et d'authentification. Leur dénominateur commun, est le besoin d'un moyen simple, pratique, fiable, pour vérifier l'identité d'une personne, sans l'assistance d'une autre personne. Le marché du contrôle d'accès s'est ouvert avec la prolifération de systèmes, mais aucun ne se révèle efficace contre la fraude, car tous utilisent un identifiant externe tel que : badge/carte, clé, code...

Il est fréquent d'oublier le code d'accès. Pour éviter cet oubli, beaucoup de personnes écrivent ce code sur un carnet, perdant ainsi toute confidentialité. Les moyens biométriques, permettent une authentification sûre, ce qui n'est pas le cas avec les mots de passe ou les cartes (badges). Ces derniers peuvent être utilisés par des tiers non autorisés.

Le niveau de sécurité d'un système est toujours celui du maillon le plus faible. Ce maillon faible, c'est bien souvent l'être humain : mot de passe aisément déchiffrable ou noté à côté de l'ordinateur. Dans la plupart des entreprises, on exige que les mots de passe soient modifiés régulièrement et comportent au moins 8 caractères, mélangeant lettres majuscules, minuscules et chiffres. L'objectif est d'échapper aux logiciels de décodage qui peuvent en peu de temps, balayer tous les mots du dictionnaire. Une protection qui peut s'avérer insuffisante pour l'accès à des applications sensibles.

Le défaut commun à tous les systèmes d'authentification est que l'on identifie un objet (ordinateur, carte, code...) et non la personne elle-même. Il est pourtant plus acceptable d'authentifier une personne, plutôt qu'une machine [biométrie].

Les technologies biométriques de reconnaissance apportent la simplicité et le confort aux utilisateurs et un niveau de sécurité jamais atteint, tout en étant superposables avec les systèmes classiques existants. Elles procurent une ergonomie non négligeable dans leur utilisation et sont une brique dans tout système de sécurité actuel et futur. Cette technologie est applicable à un large champ d'applications (contrôle d'accès, gestion horaire, paiement sécurisé sur Internet, login sur ordinateur, etc.).

### **1.2. Un peu d'histoire [biométrie]:**

Le procédé de reconnaissance anthropométrique le plus ancien, bien connu du grand public, est l'analyse des empreintes digitales, ce qui explique aujourd'hui son importance parmi les nombreux procédés d'identification sur la base de caractères physiologiques. Ce type d'analyse est aujourd'hui très au point, mais la recherche à son sujet ne date pas d'hier.

Bien que les empreintes de mains laissées par nos ancêtres sur les parois des cavernes, ne nous aient pas livré tous leurs secrets, l'empreinte du pouce servait déjà de signature lors d'échanges commerciaux à Babylone (-3000 av. JC) et dans la Chine

antique (7ème siècle). Plus près de chez nous, Marcello Malpighi, médecin et biologiste italien, décrit les composantes de nombreux organes comme le foie et les reins, recherches très critiquées à l'époque. Il décrit la couche basale de l'épiderme à laquelle il donna son nom et étudia dès le 17ème siècle les dessins et les courbes des empreintes. Le physiologiste tchèque Jan Evangelista Purkinje, précurseur de l'embryologie et de la pharmacologie, travailla sur différentes parties du corps humain dont les empreintes digitales et publia aux alentours de 1820 une thèse dans laquelle il étudiait différents types d'empreintes .

Ce n'est pourtant qu'à partir de 1870 que les empreintes digitales furent réellement utilisées à des fins d'identification par le docteur Henry Faulds, alors chirurgien dans un hôpital de Tokyo. Il publia un article dans la revue « Nature », où il développait l'idée d'utiliser les traces de doigt pour identifier des personnes. Faulds fut par la suite le premier à établir une classification grâce à de l'encre d'imprimerie et en informa Francis Galton. Ce scientifique britannique, réputé pour ses travaux en anthropologie et dans le domaine de l'hérédité, était passionné par la théorie de l'évolution, bien qu'il s'intéressait à de nombreux autres domaines scientifiques comme la météorologie ou les statistiques. Il réalisa d'importants travaux sur les mensurations des êtres humains, et rassembla des statistiques sur la taille, le poids, les dimensions et d'autres caractères de nombreuses personnes. Galton consacra une attention particulière aux empreintes digitales et mit au point une méthode d'identification dans son ouvrage « Fingerprints », où il établissait l'unicité et la permanence des figures cutanées. Son approche globale des populations lui permit de constater l'importante variation des empreintes dans l'espèce humaine mais aussi à l'intérieur de différentes ethnies. Il démontra également qu'il y avait seulement une chance sur 64 milliards que deux individus aient les mêmes figures digitales et affirma la pérennité de celles-ci chez un même individu au cours de sa vie. Grâce à lui, au Bengale, et ce dès 1858, l'empreinte du pouce servait de preuve d'identité .

Les premiers prototypes de terminaux de traitement automatique d'empreintes digitales arrivent dans le milieu des années 1970, et les premiers produits commerciaux mis sur le marché au début des années 1980. Ces terminaux seront utilisés dans un premier temps pour du contrôle d'accès et/ou de gestion du temps pour clients du type "organisation gouvernementale" (prisons par exemple).

### **1.3. Qu'est ce que la Biométrie [biométrie]?**

#### **3 possibilités pour prouver son identité :**

1. Ce que l'on possède (carte, badge, document) ;
2. Ce que l'on sait (un nom, un mot de passe) ;
3. Ce que l'on est (empreintes digitales, main, visage...) - Il s'agit de la biométrie.

#### **Il existe deux types d'accès.**

- L'accès physique qui désigne tout ce qui est physiquement accessible comme un bâtiment, une salle de laboratoire, un bureau...
- L'accès logique qui désigne tout ce qui est virtuellement accessible comme un site Internet, un fichier informatique, une application informatique...

La biométrie permet l'identification ou l'authentification d'une personne sur les bases de données reconnaissables et vérifiables qui lui sont propres.

### **3 catégories de technologies biométriques :**

1. Analyses biologiques : Odeur, sang, salive, urine, ADN...
2. Analyses comportementales : La dynamique de la signature (la vitesse de déplacement du stylo, les accélérations, la pression exercée, l'inclinaison), la façon d'utiliser un clavier d'ordinateur (la pression exercée, la vitesse de frappe), la voix, la manière de marcher (démarche)...
3. Analyses morphologiques : empreintes digitales, forme de la main, traits du visage, dessin du réseau veineux de l'œil.... Ces éléments ont l'avantage d'être stables dans la vie d'un individu et ne subissent pas autant les effets du stress par exemple, que l'on retrouve dans l'identification comportementale.

### **2 modes de fonctionnement :**

1. Authentification (1:N): dans ce mode, on pose la question : « suis-je bien M. X ? ». Techniquement, le dispositif vérifie par rapport à un code (identifiant) saisi sur un clavier, ou lu par le passage d'un badge (carte à puce, magnétique, proximité, etc.) que l'échantillon biométrique fourni correspond bien au gabarit désigné par l'identifiant.
2. Identification (1:1): pour ce mode, on pose la simple question : « qui suis-je ? ». A partir de l'échantillon biométrique fourni, le dispositif recherche le gabarit correspondant dans sa base de données.

### **Les caractéristiques collectées doivent être :**

1. universelles (exister chez tous les individus),
2. uniques (permettre de différencier un individu par rapport à un autre),
3. permanentes (autoriser l'évolution dans le temps),
4. enregistrables (collecter les caractéristiques d'un individu avec l'accord de celui-ci),
5. mesurables (autoriser une comparaison future).

#### **1.4. Pourquoi utiliser la biométrie [biométrie]?**

La biométrie est un domaine émergent où la technologie améliore notre capacité à identifier une personne. La protection des consommateurs contre la fraude ou le vol est un des buts de la biométrie. L'avantage de l'identification biométrique est que chaque individu a ses propres caractéristiques physiques qui ne peuvent être changées, perdues ou volées. La méthode d'identification biométrique peut aussi être utilisée en complément ou remplacement de mots de passe.

Plusieurs raisons peuvent motiver l'usage de la biométrie :

- Une haute sécurité - En l'associant à d'autres technologies comme le cryptage, la carte à puce...
- Confort - En remplaçant juste le mot de passe, exemple pour l'ouverture d'un système d'exploitation, la biométrie permet de respecter les règles de base de la sécurité (ne pas inscrire son mot de passe à côté du PC, ne pas désactiver

l'écran de veille pour éviter des saisies de mots de passe fréquentes). Et quand ces règles sont respectées, la biométrie évite aux administrateurs de réseaux d'avoir à répondre aux nombreux appels pour perte de mot de passe (que l'on donne parfois au téléphone, donc sans sécurité).

- Sécurité / Psychologie - Dans certains cas, particulièrement pour le commerce électronique, l'utilisateur n'a pas confiance. Il est important pour les acteurs de ce marché de convaincre le consommateur de faire des transactions. Un moyen d'authentification connu comme les empreintes digitales pourrait faire changer le comportement des consommateurs.

## **1.5. Les systèmes biométriques [biométrie]:**

### **1.5.1. Les empreintes :**

L'empreinte digitale est le modèle du relief cutané des doigts. Ce relief se forme durant la période foetale. Ses propriétés biologiques sont bien comprises. Employée pendant des siècles, sa validité d'identification est bien établie .

L'utilisation de l'empreinte digitale comme moyen d'identification d'une personne n'est pas nouvelle. En fait, les corps policiers utilisent cette technique depuis plus de 100 ans. Aujourd'hui, les empreintes digitales sont recueillies sur une scène de crime et sont ensuite comparées à celles contenues dans un serveur central.

Le recours à l'empreinte digitale compte pour plus du tiers du marché des procédés biométriques. Elle représente nettement la solution préférée des entreprises œuvrant dans ce domaine. La force de ce procédé tient au fait que l'utilisation de l'empreinte digitale est généralement plus facile d'acceptation par la communauté et qu'elle est une des plus efficaces et des moins coûteuses.

On classe les empreintes selon un système vieux d'une décennie : le système Henry. Dans ce système, le classement repose sur la topographie générale de l'empreinte digitale et permet de définir des familles (figure 1.1) telles que les boucles (à gauche ou à droite), les arches et les tourbillons.

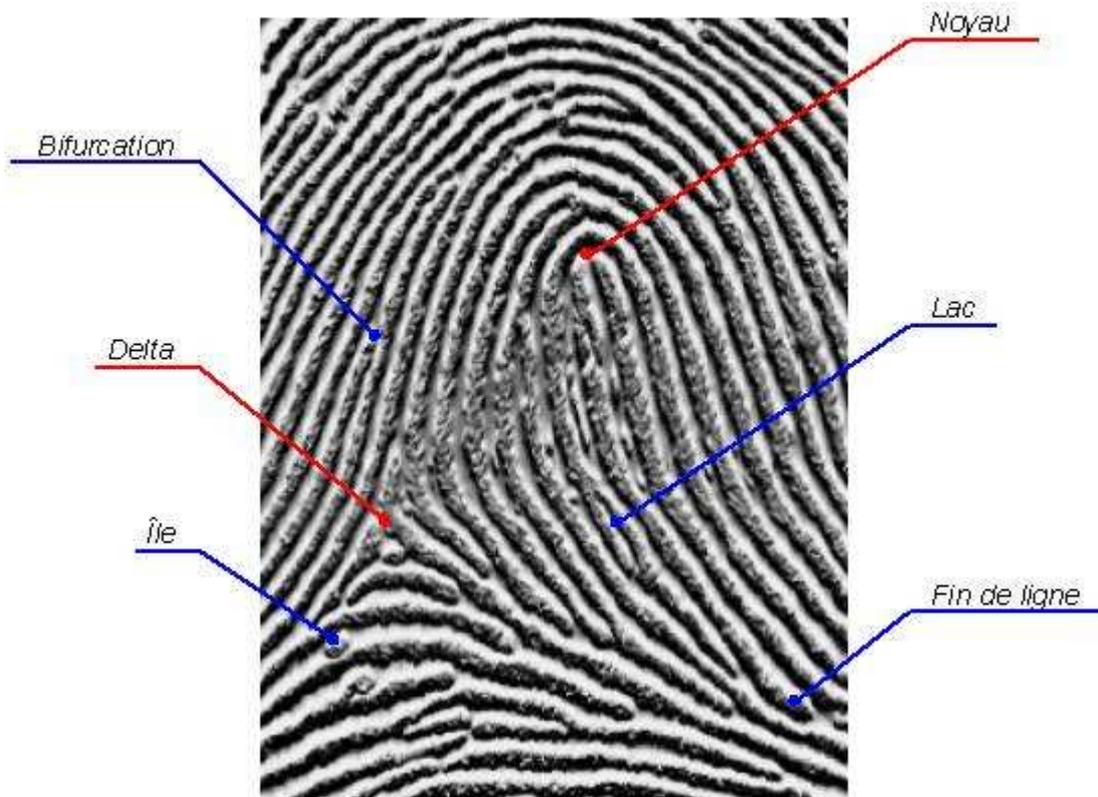


**Figure 1.1** Topographie générale de l'empreinte digitale.

Les éléments qui permettent de différencier deux empreintes digitales sont les minuties. Une minutie est un point qui se situe sur le changement de continuité des lignes papillaires.

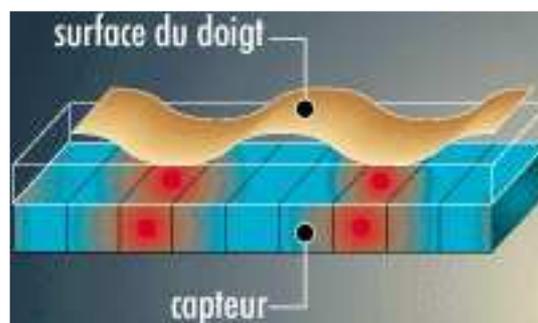
La probabilité de trouver deux empreintes digitales similaires est de 1 sur 10 puissance 24. Dans la pratique, il est évident que deux jumeaux issus de la même cellule ont des empreintes digitales très proches.

La figure 2.1 montre les caractéristiques topographiques d'une empreinte.



**Figure 1.2** Caractéristiques topographiques d'une empreinte.

La capture de l'image d'une empreinte digitale consiste à trouver les lignes tracées par les crêtes (en contact avec le capteur) et les vallées (creux). Le point commun à toutes les technologies utilisées pour la prise d'image d'une empreinte, est que l'image est constituée à partir des points de contact du doigt sur le capteur (voir figure 1.3).



**Figure 1.3** Capture de l'image d'une empreinte en se basant sur les points de contact du doigt sur le capteur.

Les types de capteurs utilisés sont:

- Le capteur optique.
- Le capteur en Silicium.
- Le capteur ultrasonique.

### **1.5.2. La main :**

La biométrie par la forme de la main est une technologie populaire qui est largement employée pour le contrôle d'accès physique ou le pointage horaire. Elle est très simple et bon marché. L'exactitude d'un système biométrique basé sur la forme de la main est tout à fait raisonnable. Les éléments pris en compte ne reposent que sur la géométrie de la main et non sur l'empreinte palmaire.

Le système prend une photo de la main et examine 90 caractéristiques, y compris la forme tridimensionnelle de la main, de la longueur et de la largeur des doigts et de la forme des articulations.

Pour utiliser la géométrie de la main, l'utilisateur place sa main sur une platine possédant des guides pour positionner les doigts.

Les lecteurs de la géométrie de main (figure 1.4) offrent un niveau très raisonnable d'exactitude, mais peuvent avoir des taux de fausses acceptations élevées pour des jumeaux ou d'autres membres de la même famille.



**Figure 1.4** Lecteur de la géométrie de la main.

### **1.5.3. La voix :**

L'identification de la voix est considérée par les utilisateurs comme une des formes les plus normales de la technologie biométrique, car elle n'est pas intrusive et n'exige aucun contact physique avec le lecteur du système.

La technologie d'analyse de la voix (aussi appelée analyse du locuteur) s'applique avec succès là où les autres technologies sont difficiles à employer. Elle est utilisée dans des secteurs comme les centres d'appel, les opérations bancaires, l'accès à des comptes, sur PC domestiques, pour l'accès à un réseau ou encore pour des applications judiciaires.

La plupart des systèmes d'identification de la voix utilisent l'affichage d'un texte, des mots spécifiques doivent être lus puis parlés afin de vérifier que la personne à authentifier est bien présente et qu'il ne s'agit pas d'un enregistrement.

Les systèmes d'identification de la voix se concentrent sur les seules caractéristiques de voix qui sont uniques à la configuration de la parole d'un individu. Ces configurations de la parole sont constituées par une combinaison des facteurs comportementaux et physiologiques.

Les imitateurs essayent habituellement de reproduire les caractéristiques vocales qui sont les plus évidentes au système auditif humain et ne recréent pas les caractéristiques moins accessibles qu'un système automatisé d'identification de voix analyse. Il n'est donc pas possible d'imiter la voix d'une personne inscrite dans la base de données.

#### **1.5.4. La signature :**

Signer un document pour s'identifier est un geste naturel. Que ce soit pour confirmer une transaction sur son compte de carte de crédit ou simplement pour apposer sa marque sur une carte d'anniversaire, nous sommes tous habitués à tracer notre griffe.

Chaque personne a un style d'écriture unique. On peut donc définir, à partir de la signature d'une personne, un modèle qui pourra être employé pour effectuer une identification. De plus, la signature est utilisée dans beaucoup de pays comme élément juridique ou administratif. Elle permet de justifier de la bonne foi d'une personne ou de la confondre devant des documents signés.

La vérification (dynamique) de signature est basée sur la différenciation entre les parties d'une signature qui sont habituelles et celles qui changent avec presque chaque signature.

Les systèmes d'authentification de signature incluent habituellement un crayon lecteur et une tablette à digitaliser. La vérification est accomplie en analysant un certain nombre de variables, dont la vitesse avec laquelle la signature est réalisée, les accélérations, la pression exercée ...



**Figure 1.5** Dispositif de vérification de signature.

Les difficultés liées à la capture d'une signature viennent du fait qu'une personne ne signe jamais deux fois de la même façon, même à quelques secondes d'intervalle. En effet suivant les émotions ou la fatigue, une signature peut fortement évoluer. D'où la mise au point d'algorithmes très complexes capables de prendre en compte ces évolutions possibles.

### **1.5.5. Le clavier :**

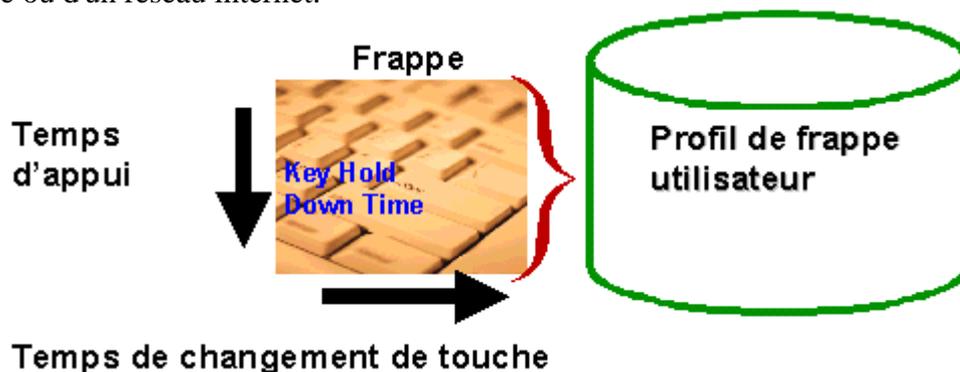
Il s'agit d'une technique de reconnaissance des personnes basée sur le rythme de frappe qui leur est propre. C'est une solution biométrique « Software Only ». Elle est appliquée au mot de passe qui devient ainsi beaucoup plus difficile à « imiter ».

Lors de la mise en place de cette technique, il est demandé à l'utilisateur de saisir son mot de passe une dizaine de fois de suite. A l'aide d'un algorithme qui exploite le temps d'appui sur chaque touche et le temps entre chaque touche, la dizaine de saisie est « moyennée » pour bâtir un Profil de frappe » de l'utilisateur qui servira de référence.

Aux accès suivants, en suivant le même approche, la saisie du mot de passe donnera sera couplée à un profil de frappe qui sera comparé au profil de référence. Le droit d'accès est alors accordé en fonction du niveau de ressemblance de ce profil avec la référence. Suivant le degré de filtrage qu'un administrateur aura défini, cet accès sera plus ou moins difficile.

Que ce soit en sécurité additionnelle à une application, à un Single Sign On, en Intranet ou en Internet, en complément de sécurité à une carte à puce personnelle, c'est une solution, fiable, simple à mettre en œuvre et très compétitive car sans Hardware.

Plus il y a d'utilisateurs, plus cette simplicité est un atout majeur. Elle se distribue aussi aisément aux quelques utilisateurs d'une PME qu'à ceux d'une grosse société ou d'un réseau internet.



**Figure 1.6** Procédure d'authentification par clavier.

### **1.5.6. L'iris:**

L'utilisation de l'iris pour l'identification d'une personne a été proposée à l'origine en 1936 par l'ophtalmologue *Frank Burch*. En 1987 les ophtalmologues *Aran Safir* et *Léonard Flom* ont fait breveter cette idée et en 1989 ils ont demandé à *John Daugman* (alors enseignant à l'université de Harvard), d'essayer de créer des algorithmes pour l'identification par l'iris. Ces algorithmes (méthode basée sur les ondes de Gabo), que *Daugman* a fait breveter en 1994, sont la base de tous les systèmes d'identification par l'iris.

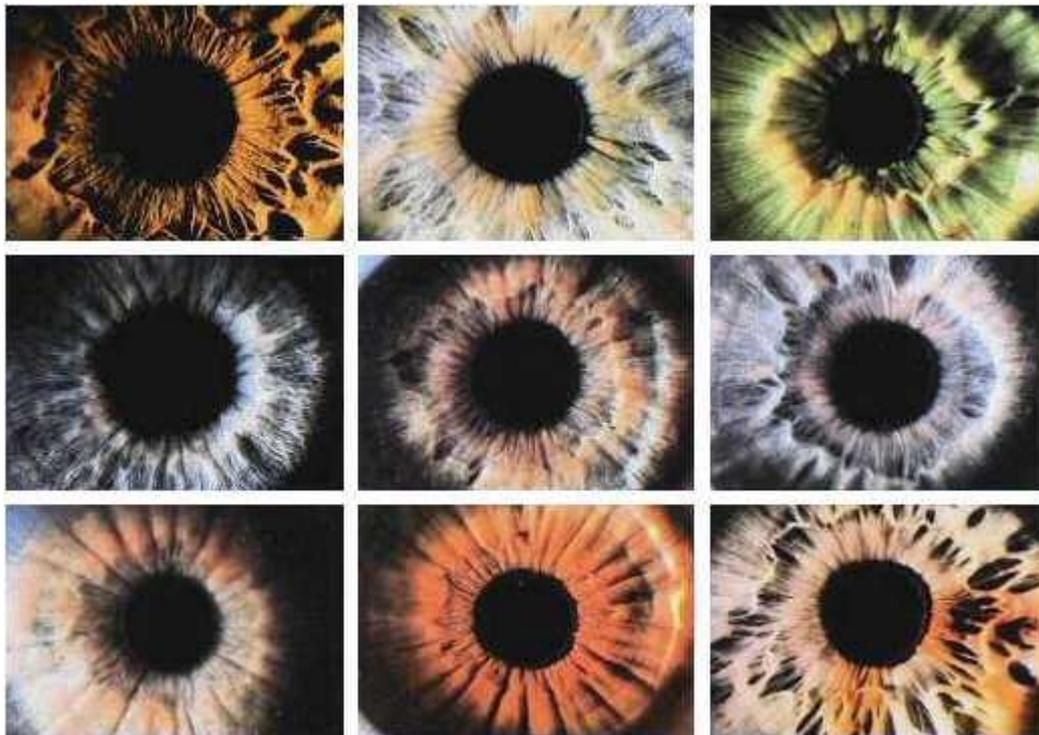
L'identification par l'iris utilise plus de paramètres que les autres méthodes d'identification et la fiabilité résultante est suffisante pour ne plus faire de l'identification mais de l'authentification.

La probabilité de trouver 2 iris suffisamment identiques est 1 sur 10 puissance 72 (selon les estimations de *Daugman*).

Deux vrais jumeaux ont assez d'éléments distinctifs sur leur iris permettant de les distinguer l'un de l'autre.

Cette formation est chaotique, elle génère donc des motifs possédant une forte variabilité (244 degrés de liberté).

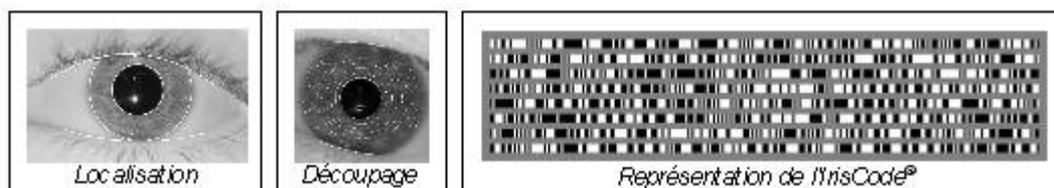
La figure 1.7 montre des formes de l'iris.



**Figure 1.7** Différentes formes de l'iris de l'oeil humain.

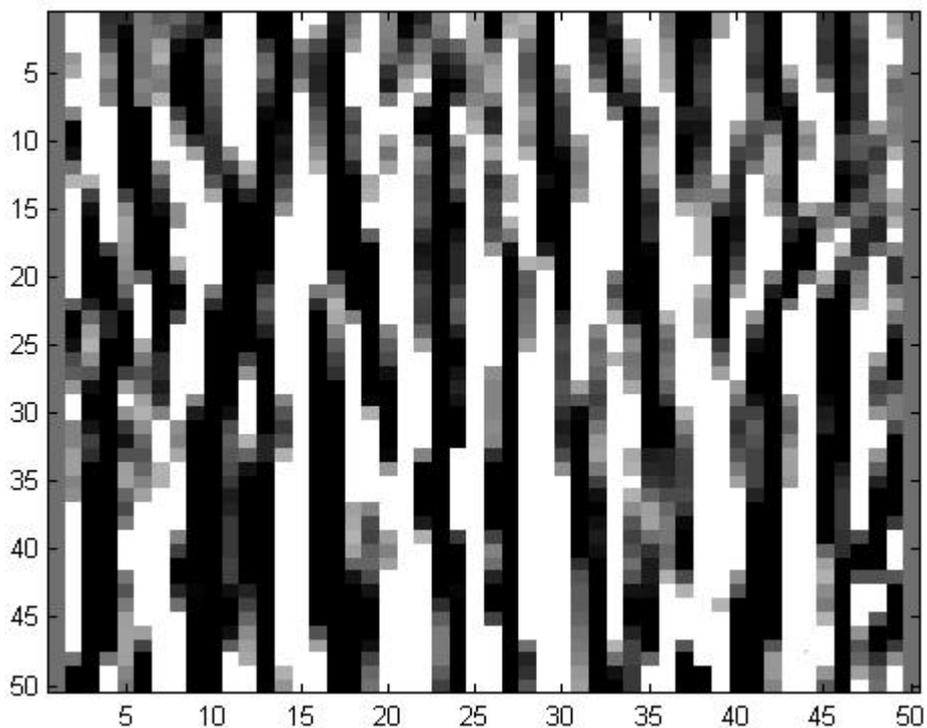
Pour distinguer les iris, on utilise les sillons de contraction, les cryptes, les anneaux, etc.

La méthode employée pour la caractérisation d'un iris (pour le moment) est celle brevetée par *John Daugman*. Après la numérisation de l'image de l'oeil, le logiciel va déterminer le centre de la pupille et la zone où se trouve l'iris. Puis sur deux secteurs angulaires, le logiciel y découpe des bandes de taille régulière et en génère un gabarit (IrisCode®) à partir de l'analyse locale de la texture de l'iris (voir figure 1.8).



**Figure 1.8** Traitement de l'image de l'iris.

La seconde étape consiste à extraire les paramètres caractéristiques de l'iris. La plupart du temps, on utilise une transformée par ondelettes comme montré dans la figure 1.9.

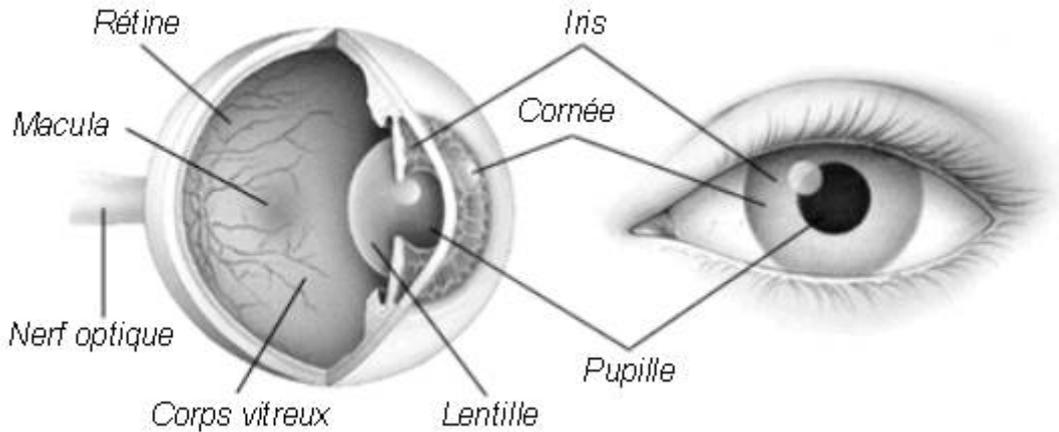


**Figure 1.9.** Transformée par ondelettes d'une image de l'iris.

### 1.5.7. La rétine:

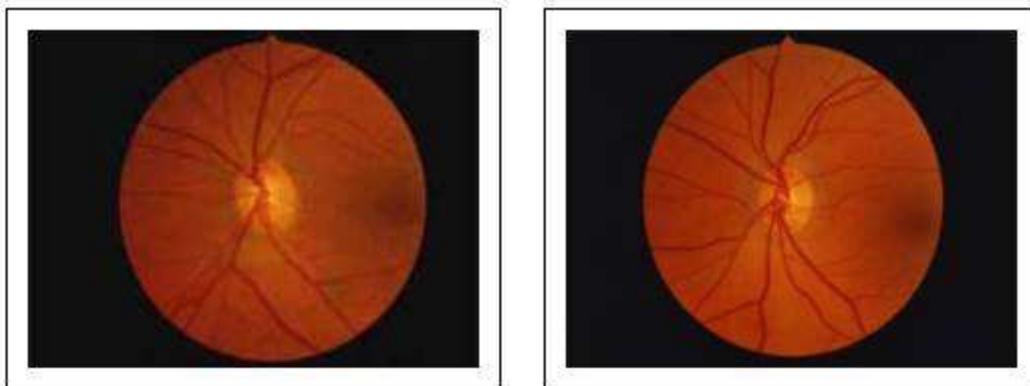
La rétine est la « pellicule photographique » de l'œil. Elle est constituée de 4 couches de cellules et est située on fond de l'œil.

La figure 1.10 présente l'anatomie de l'œil humain.



**Figure 1.10** Anatomie de l'œil humain.

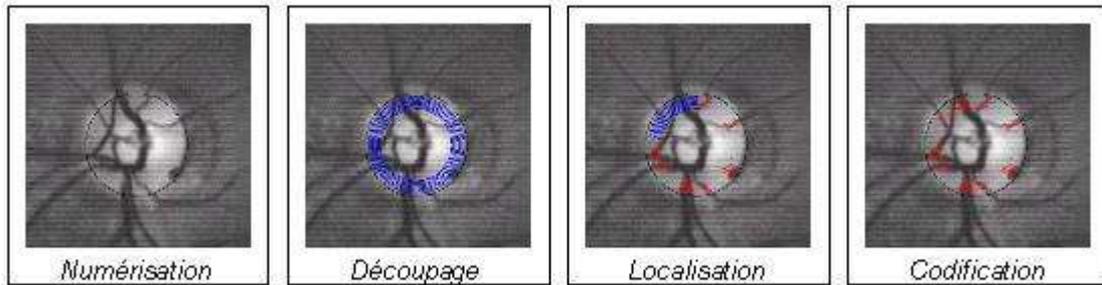
Les éléments qui permettent de distinguer deux rétines sont les veines qui les tapissent. La disposition de ces veines est stable et unique d'un individu à l'autre d'un œil à l'autre). Et les modèles qui en sont issus héritent de la stabilité de cette disposition (voir figure 1.11).



**Figure 1.11** Photographie des deux rétines d'un individu.

De plus, pour un fraudeur, il est impossible de pouvoir reproduire ce motif. D'où un très haut niveau de sécurité.

Après la capture d'une image de la rétine, le logiciel du dispositif de lecture découpe un anneau autour de la fovéa. Dans cet anneau il repère l'emplacement des veines et leur orientation. Puis il les codifie dans un gabarit (voir figure 1.12). L'opération en elle même est assez simple à décrire mais les algorithmes restent relativement complexes.



**Figure 1.12** Caractérisation d'une rétine

### **1.5.8. Le visage:**

Rien n'est plus naturel qu'utiliser le visage pour identifier une personne. Les images faciales sont probablement la caractéristique biométrique la plus communément employée par l'homme pour effectuer une identification personnelle.

L'utilisation d'une caméra permet de capter la forme du visage d'un individu et d'en dégager certaines particularités. Selon le système utilisé, l'individu doit être positionné devant l'appareil ou peut être en mouvement à une certaine distance. Les données biométriques qui sont obtenues sont par la suite comparées au fichier référence.

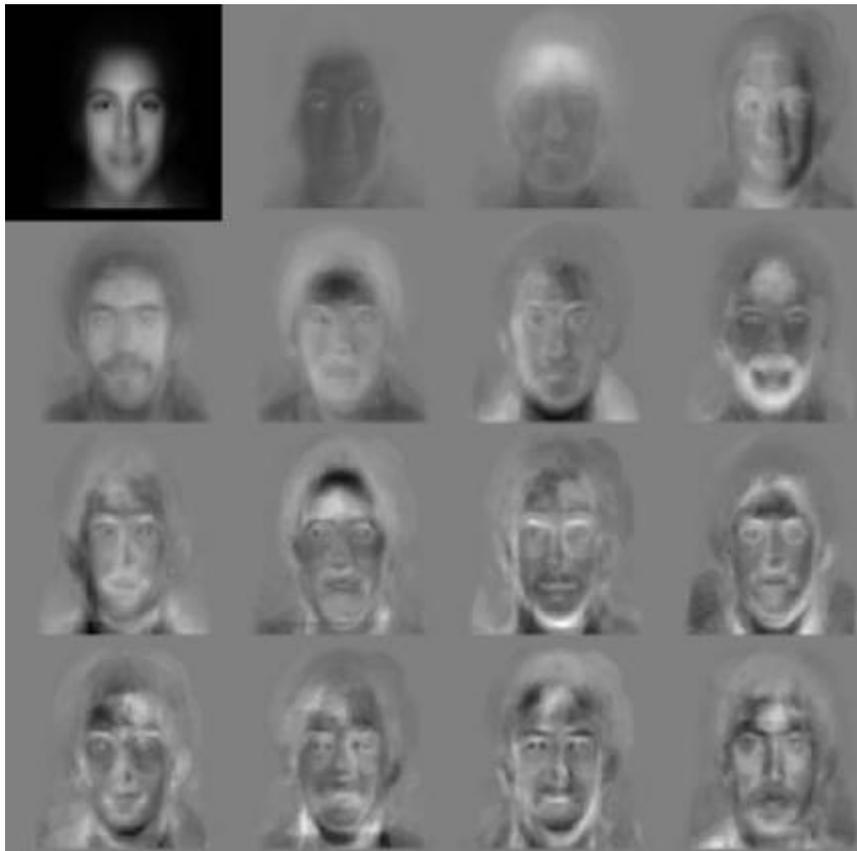
Au début des années 1970, la reconnaissance par le visage était principalement basée sur des attributs faciaux mesurables comme l'écartement des yeux, des sourcils, des lèvres, la position du menton, la forme, etc. Depuis les années 1990, les différentes technologies utilisées exploitent toutes les découvertes effectuées dans le domaine du traitement d'image, et plus récemment encore les réseaux de neurones.

Il existe plusieurs techniques de reconnaissance par analyse du visage. Mais pour la plupart il est d'intérêt que ces techniques se basent sur des éléments du visage qui sont le moins susceptibles aux changements : les grands traits supérieurs des orbites, les secteurs entourant les pommettes, les côtés de la bouche et d'autres caractéristiques similaires de façon à ignorer les changements comme la coupe de cheveux. Toutes ces techniques conduisent à des fonctionnements en identification sur des bases de milliers voir de centaines de milliers de personnes.

Quatre méthodes de base sont utilisées par les fabricants de systèmes biométriques : Eigenfaces, l'analyse de points particuliers, les réseaux de neurones et le traitement automatique de visage.

- Le traitement automatique de visage est une technologie rudimentaire, elle caractérise les visages par des distances et des proportions entre des points particuliers comme les yeux, le nez les coins de la bouche. Aussi éprouvé que les autres technologies, le traitement automatique de visage est la plus efficace dans des situations de capture d'image avec peu d'éclairage.
- Eigenfaces : elle utilise une représentation des éléments caractéristiques d'une image de visage à partir d'images propres en niveau de gris (voir figure 1.13). Des variantes des Eigenfaces sont fréquemment utilisées comme base pour d'autres méthodes de reconnaissance.

- L'analyse de points particuliers est la technologie d'identification faciale la plus largement utilisée. Cette technologie se rapproche des Eigenfaces, mais elle est capable de s'adapter à des changements d'aspect facial (sourire, froncer des sourcils,...).
- Les réseaux de neurones permettent en théorie de mener à une capacité accrue d'identification dans des conditions difficiles de capture. Les réseaux de neurones emploient un algorithme pour déterminer la similitude entre des captures d'images de visage, et des gabarits. Ceci en employant le plus possible de captures et de gabarits.

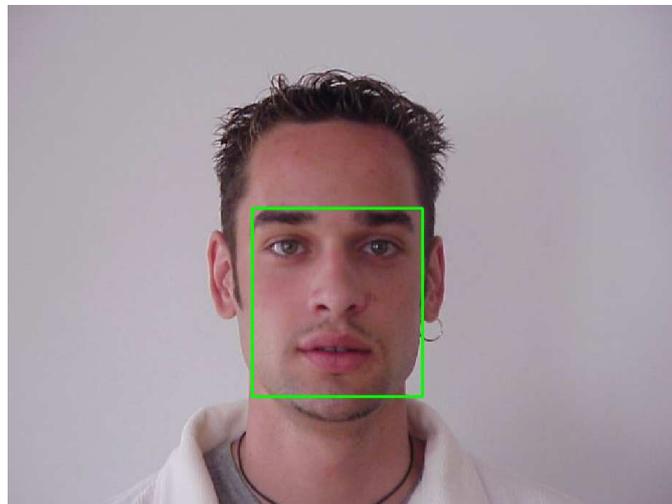


**Figure 1.13** La méthode des Eigenfaces (visages propres), source MIT.

### **1.6. Conclusion:**

On peut constater que la biométrie est une véritable alternative aux mots de passe et autres identifiants. Elle permet de vérifier que l'utilisateur est bien la personne qu'il prétend être. Cette technologie est en pleine croissance et tend à s'associer à d'autres technologies comme la carte à puce

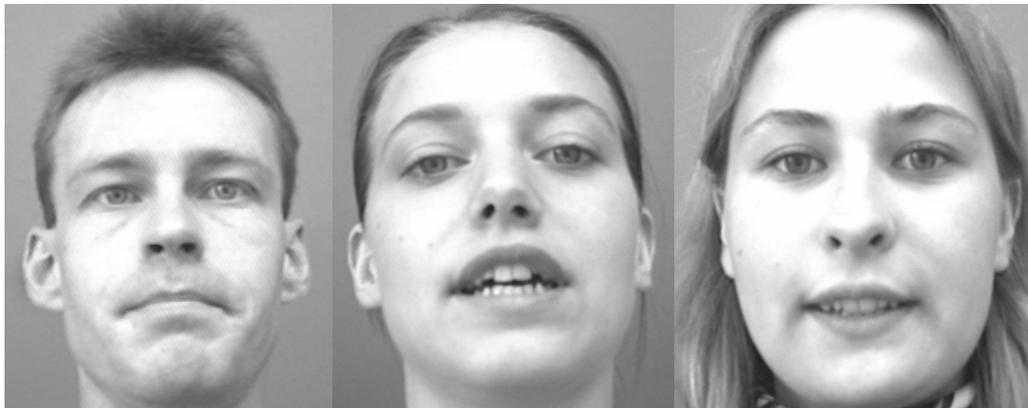
# Chapitre 2: La détection de visage



### **2.1. Introduction:**

Les derniers progrès dans les technologies des ordinateurs ont permis d'améliorer la vie des humains grâce à l'intelligence artificielle. En effet, cette tendance a déjà incité un développement actif dans l'intelligence des machines. La vision par ordinateur, par exemple, vise à reproduire la vision humaine. Traditionnellement, des systèmes de vision par ordinateur ont été utilisés dans des tâches spécifiques telles que les tâches visuelles répétitives et fastidieuses d'inspection dans des chaînes de montage. Le développement actuel dans ce domaine se dirige vers des applications de vision plus généralisées, telles que la reconnaissance de visage et les techniques de codage des vidéos.

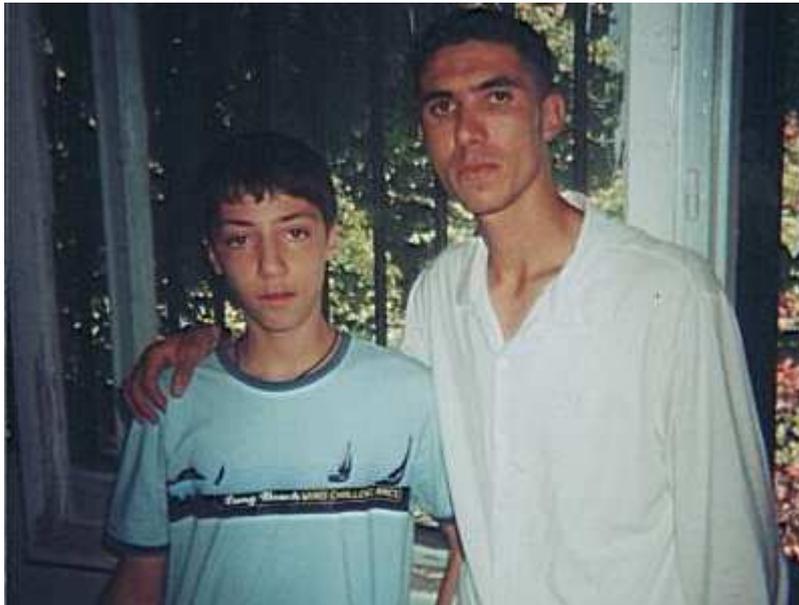
Plusieurs techniques actuelles de reconnaissance de visage supposent la disponibilité de visages de vue frontale et de taille égale [Chellappa et autres 1995 ; Samal et Iyengar 1992]. En réalité, cette condition ne peut pas être satisfaite à cause de la nature variable des apparences des visages et des conditions d'environnement. Considérons les images de la figure 2.1 (tirées de la base de données des visages *BioID face database*). Ces images sont des images de test typiques utilisées dans la recherche en classification de visage. L'exclusion de l'arrière plan dans ces images est une étape nécessaire pour des techniques fiables de classification de visage. Cependant, dans des scénarios d'applications réelles comme dans l'exemple montré dans la figure 2.2, un visage peut apparaître dans un arrière plan complexe et à différentes positions. Les systèmes de reconnaissance qui sont basés sur des images standard de visage sont susceptibles de confondre quelques régions de l'arrière plan avec des visages. Afin de rectifier le problème, un processeur visuel d'entrée est nécessaire pour localiser et extraire la région de visage à partir de l'arrière plan.



**Figure 2.1** Images typiques pour la reconnaissance du visage.

La détection de visage est une des tâches visuelles que les humains peuvent accomplir sans effort. Cependant, en termes de vision par ordinateur, cette tâche n'est pas facile. Une formulation générale du problème peut être définie comme suit: Etant donnée une image statique (still image) ou une vidéo (video image), détecter et localiser un nombre inconnu (s'il existe) de visages. La solution à ce problème implique la segmentation, l'extraction, et la vérification des visages et probablement des traits faciaux à partir d'un arrière plan non contrôlé. Comme processeur visuel d'entrée, un système de détection de visage devrait également pouvoir réaliser la tâche indépendamment de l'illumination, de l'orientation, et de la distance de la camera. Ce chapitre vise à fournir un aperçu de la recherche contemporaine en détection de visage

et d'une manière structurée. [Chellappa et autres 1995] ont conduit une étude détaillée sur la recherche en reconnaissance de visage. Dans leur étude, plusieurs aspects, y compris la segmentation et l'extraction des traits, liées à la reconnaissance de visage ont été passés en revue. Une des conclusions de [Chellappa et autres 1995] était que le problème de détection de visage a suscité étonnamment peu d'attention. Ceci a certainement changé au cours des dernières années comme on le montrera dans cet aperçu.



**Figure 2.2** Un scénario réel de détection de visages.

### **2.2. L'évolution de la recherche sur la détection de visage:**

Les premiers efforts dans la détection de visage remontent au début des années 1970, où de simples techniques heuristiques et anthropométriques [Sakai et autres 1972] étaient utilisées. Ces techniques étaient largement rigides à cause des diverses conditions imposées comme l'arrière plan uni, la vue frontale du visage (une photographie typique d'un passeport),... Pour ces systèmes, le moindre changement de ces conditions entraînait un réglage délicat sinon un design total du système. En dépit de ces problèmes, l'évolution de l'intérêt de la recherche s'est stagnée jusqu'aux années 1990 [Chellappa et autres 1995], où les systèmes pratiques de reconnaissance de visage et de codage de vidéo commencent à voir le jour. Au cours de la dernière décennie, il y avait un grand nombre de recherches qui couvraient plusieurs aspects importants de la détection de visage. Un grand nombre d'algorithmes robustes pour la segmentation ont été présentés, particulièrement ceux utilisant le mouvement, la couleur et l'information généralisée. L'utilisation des statistiques et des réseaux de neurones a permis aussi la détection de visage à partir d'environnements complexes et se trouvant à différentes distances de la caméra. En plus, il y a de nombreuses avancés dans la conception d'extracteurs de traits tels que les *modèles déformables* et les *contours actifs* qui peuvent localiser et dépister les traits faciaux avec exactitude.

Puisque les techniques de détection de visage exigent *l'information a priori* du visage, elles peuvent être efficacement organisées en deux grandes catégories distinguées par leur approche différente à utiliser la connaissance de visage. Les techniques de la première catégorie utilisent explicitement la connaissance du visage

et suivent la méthodologie classique de détection dans laquelle des traits de bas niveau sont dérivés avant l'analyse basée sur la connaissance [Brunelli et Poggio 1993 ; Valentin et autres 1994]. Les propriétés apparentes du visage telles que la géométrie du visage et la couleur de la peau sont exploitées à différents niveaux du système. Typiquement, dans ces techniques, les tâches de détection de visage sont accomplies par la manipulation des mesures de distance, des angles et de surface des traits visuels dérivés de la scène. Puisque les traits constituent les ingrédients principaux, cette catégorie de techniques est appelée *l'approche basée sur les traits (feature-based approach)*. Ces approches ont incarné la majorité d'intérêt pour la recherche en détection de visage commençant dès les années 70. Profitant de l'avantage fourni par l'évolution actuelle de la théorie de la reconnaissance des formes, les techniques de la deuxième catégorie considèrent la détection de visage comme un problème général de reconnaissance de formes. Les représentations de visage basées sur l'image (image-based representations) [Demirel et autres 1996], par exemple dans des tableaux d'intensité lumineuse à deux dimensions, sont directement classées en groupe de visage moyennant des algorithmes d'apprentissage et sans extraction ou analyse de traits. Cette approche est appelée *l'approche basée sur l'image (image-based approach)*. Contrairement à l'approche basée sur les traits, cette approche relativement nouvelle incorpore implicitement [Valentin et autres 1994] l'information du visage dans le système de détection, à travers des schèmes d'apprentissage et de projection.

La figure 2.3 donne une idée des approches de détection de visage et de leurs dérivées.

### **2.3. L'approche basée sur les traits:**

Le développement de l'approche basée sur les traits peut être divisé en trois secteurs. Soit donné un problème typique de détection de visage qui consiste à localiser un visage dans une scène complexe, premièrement, l'analyse de bas niveau traite la segmentation des traits visuels en utilisant les propriétés des pixels de l'image, telles que le niveau de gris et la couleur. A cause de la nature de "bas niveau" de ce traitement, les traits générés de cette analyse sont ambigus. Dans l'analyse des traits, les traits visuels sont organisés sous un concept plus global de visage et de traits de visage moyennant l'utilisation de l'information de la géométrie du visage. A travers l'analyse des traits, les ambiguïtés qui touchent les traits sont réduites et les positions du visage et de ses traits sont déterminées. Le groupe suivant implique l'utilisation des *modèles de formes actives (active shape models)*. Ces modèles qui vont des plus classiques, à savoir les *snakes*, proposés en fin des années 1980, jusqu'aux plus récents *modèles de points distribués (point distributed models ou PDM)*, ont été développés dans le but de les utiliser pour l'extraction de traits complexes et non rigides comme les pupilles des yeux et pour la poursuite de lèvres.

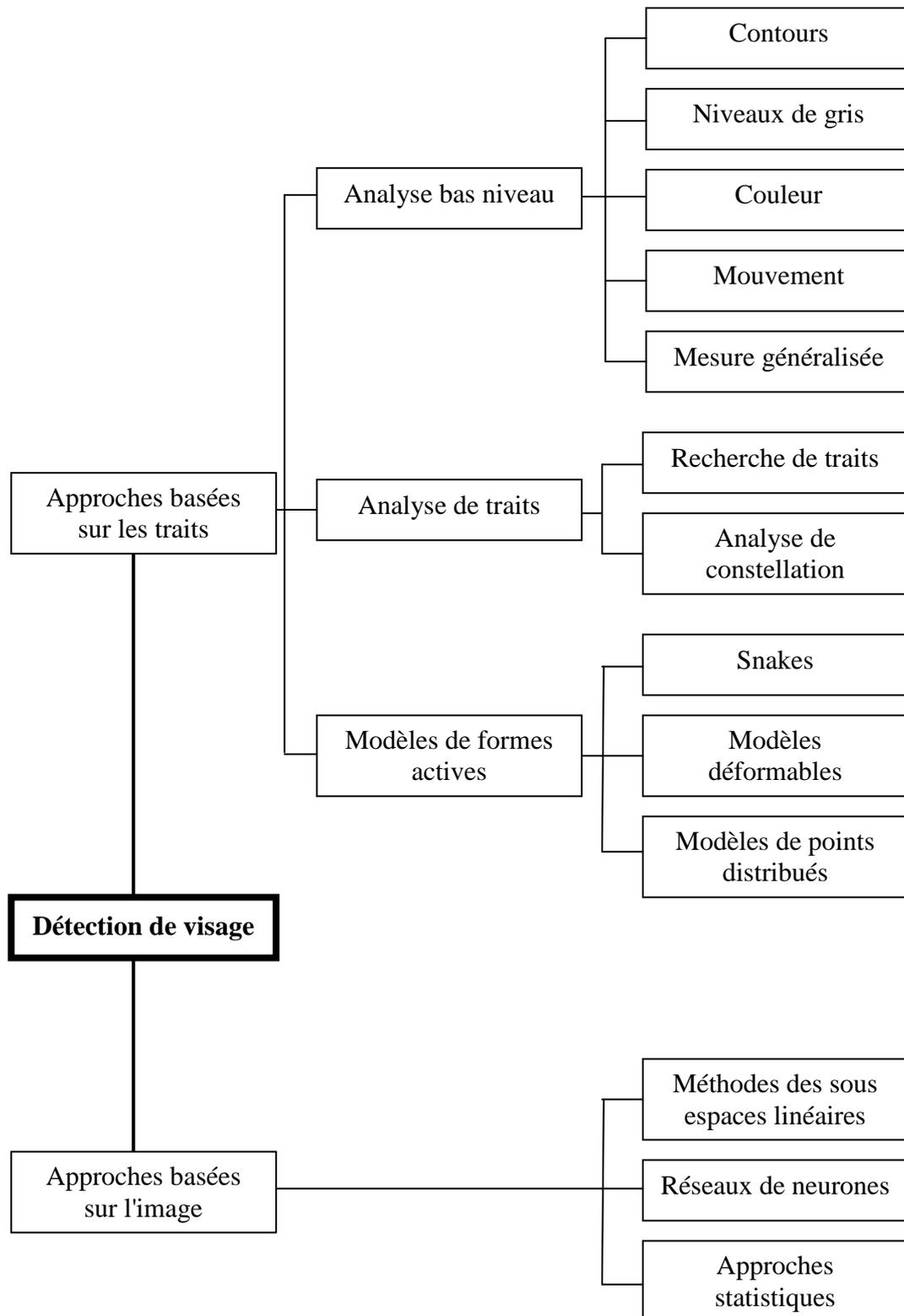


Figure 2.3 Les approches de la détection de visage.

### **2.3.1. L'analyse de bas niveau:**

#### **2.3.1.1. Les bords:**

Etant considérée comme la caractéristique la plus primitive dans les applications de vision par ordinateur, la représentation des bords a été appliquée précocement dans le travail de détection de visage de [Sakai et autres 1972]. Ce travail était basé sur l'analyse des dessins de lignes de visages à partir de photographies, ayant pour objectif la localisation des traits du visage. [Craw et autres 1987] ont proposé plus tard une méthode hiérarchique basée sur le travail de [Sakai et autres 1972] pour tracer le contour d'une tête humaine. Ce travail incluait un poursuiveur de lignes implémenté avec une contrainte de courbure pour le prévenir des éventuelles distractions causées par des bords bruités. Les traits de bords qui se trouvent sur le pourtour de la tête sont ensuite soumis à une analyse de traits en utilisant les informations de la position et de la forme du visage. Des exemples plus récents de techniques basées sur les bords peuvent être trouvés dans [Brunelli et Poggio 1993 ; Choi et autres 1999 ; Herpers et autres 1996 ; Low 1998 ; Low et Ibrahim 1997] pour l'extraction des traits de visage et dans [De Silva et autres 1995 ; Govindaraju 1996 ; Herpers et autres 1995 ; Herpers et autres 1999 ; Huang et autres 1996 ; Jacquin et Eleftheriadis 1995 ; Li et Roeder 1995 ; Wang et Tan 2000 ; Yuille et autres 1992] pour la détection de visage. Les techniques basées sur les bords ont trouvé aussi des applications dans la détection des lunettes à partir d'images faciales [Jiang et autres 2000, Jing et Mariani 2000].

La détection des bords est la toute première étape dans la dérivation de la représentation des bords. Dans la littérature, plusieurs types d'opérateurs d'extraction de bords ont été utilisés. L'opérateur de *Sobel* était le plus utilisé dans les techniques mentionnées précédemment [Brunelli et Poggio 1993 ; Choi et autres 1999 ; De Silva et autres 1995 ; Jacquin et Eleftheriadis 1995 ; Kawaguchi et autres 2000 ; Li et Roeder 1995]. L'opérateur d'extraction de bords de *Marr-Hildreth* [Marr et Hildreth 1980] fait partie des systèmes proposés dans [Govindaraju 1996 ; Huang et autres 1996]. Une variété de dérivées du premier et du second ordre (*Laplacien*) de Gaussiennes a été utilisée dans d'autres méthodes. Par exemple, un Laplacien à grande échelle a été utilisé pour obtenir un dessin de ligne dans [Sakai et autres 1972] et des filtres dirigeables et à orientation multi échelle dans [Herpers et autres 1996] et dans [Herpers et autres 1995], respectivement. Le filtrage dirigeable dans [Herpers et autres 1996] consiste en trois étapes consécutives de détection de bords qui incluent la détection des bords, la détermination de l'orientation du filtre pour chaque bord détecté et la poursuite par étapes des bords voisins en utilisant l'information de l'orientation. Cet algorithme a permis une extraction précise de plusieurs points clés de l'oeil.

Dans l'approche basée sur la détection de bords, les bords doivent être étiquetés et assortis à un modèle de visage dans le but de vérifier les détections correctes. *Govindaraju* [Govindaraju 1996] accomplissait cette tâche en étiquetant les bords en bords du côté gauche, de la ligne de naissance des cheveux et du côté droit d'une vue frontale d'un visage, il assortissait ensuite ces bords avec un modèle de visage en utilisant la *proportion magique (golden ratio)* (un rectangle proportionné esthétiquement par les artistes) d'un visage idéal:

$$\frac{\text{hauteur}}{\text{Largeur}} = \frac{1 + \sqrt{5}}{2} \quad (2.1)$$

L'extraction de traits de *Govindaraju* basée sur les bords s'effectue comme suit:

- Détection de bords: opérateur de *Marr-Hildreth*.
- Amincissement: un algorithme d'amincissement classique proposé par *Pavlidis* [Pavlidis 1982].
- Elimination d'éperons: chaque composant connecté est réduit à sa branche centrale.
- Filtrage: les composants qui n'ont pas des propriétés similaires à celles d'un visage sont éliminés.
- Détection de corners: les composants sont divisés suivant les corners détectés.
- Etiquetage: Les composants finaux sont étiquetés comme appartenant au côté gauche, à la ligne de naissance des cheveux ou au côté droit du visage.

Les composants étiquetés sont combinés pour former les positions possibles de visage et ceci en se basant sur une fonction de coût (qui utilise la proportion définie plus haut). Sur un ensemble de 60 images d'apprentissage avec arrière plan complexe et contenant 90 visages, ce système a détecté correctement 76% des visages avec une moyenne de 2 fausses alarmes par image.

### **2.3.1.2. L'information des niveaux de gris:**

Sans compter que les détails des bords, l'information des niveaux de gris dans un visage peut également être utilisée comme traits. Les traits faciaux tels que les sourcils, les pupilles et des lèvres paraissent généralement plus foncés que les régions faciales d'entourage. Cette propriété peut être exploitée pour différencier les diverses parties faciales. Plusieurs algorithmes récents d'extraction de traits faciaux [Van Beek et autres 1992 ; Graf et autres 1996 ; Lam et Yan 1994a] recherchent des minimums locaux des niveaux de gris dans des régions faciales segmentées. Dans ces algorithmes, les images d'entrée sont d'abord améliorées par augmentation du contraste et par des routines de morphologie de niveaux de gris pour améliorer la qualité des régions locales foncées et pour faciliter de ce fait la détection. L'extraction des régions foncées est réalisée par un seuillage de bas niveau des niveaux de gris. Du côté pratique, [Wong et autres 1995] ont implémenté un robot qui recherche également les régions faciales foncées chez des candidats de visage obtenus indirectement à partir de l'analyse de couleur. L'algorithme se sert d'un modèle pondéré de l'oeil humain pour déterminer les endroits possibles des yeux. Dans [Hoogenboom et Lew 1996], les maximums locaux, qui sont définis par un pixel clair entouré par huit voisins foncés sont utilisés pour indiquer les points faciaux claires (lumineux) tels que le bout du nez. Les points de détection sont alors alignés avec des modèles de traits pour des mesures de corrélation.

[Yang et Huang 1994], d'autre part, ont exploré le comportement des niveaux de gris des visages dans des images de mosaïque (pyramide). Quand la résolution d'une image de visage est réduite graduellement soit par sous échantillonnage soit en appliquant la moyenne, les traits macroscopiques du visage vont disparaître. À une basse résolution, la région de visage deviendra uniforme. Basé sur cette observation, *Yang* a proposé un cadre hiérarchique de détection de visage. Commenant par des images de basse résolution, des candidats de visage sont établis par un ensemble de règles qui recherchent des régions uniformes. Les candidats de visage sont alors vérifiés par l'existence de traits faciaux proéminents en utilisant des minimums locaux à des résolutions plus élevées. La technique de *Yang et Huang* a été récemment

incorporée à un système pour la détection invariable par rapport à la rotation par [Lv et autres 2000] et une prolongation de l'algorithme est présentée dans [Kotropoulos et Pitas 1997].

### **2.3.1.3. La couleur:**

Tandis que l'information des niveaux de gris fournit la représentation de base pour les traits d'images, la couleur est un moyen plus puissant de discerner l'apparence des objets. En raison des dimensions supplémentaires que la couleur possède, deux formes d'informations semblables des niveaux de gris pourraient apparaître très différemment dans un espace de couleur. On a constaté que les couleurs différentes de la peau humaine se regroupent dans petit cluster dans les espaces de couleur même lorsque des visages de différentes races sont considérés [Hunke et Waibel 1994 ; McKenna et autres 1996 ; Yang et Waibel 1996]. Ceci signifie que la composition de la couleur de la peau humaine diffère peu à travers les individus.

Un des modèles de couleur le plus répandus est la représentation *RGB* (*Red, Green, and Blue*) dans laquelle différentes couleurs sont définies par des combinaisons de composantes primaires de couleurs RVB (Rouge, Vert et Bleu). Puisque la variation principale de l'apparence de la peau est en grande partie due au changement de luminance (éclat) [Yang et Waibel 1996], les couleurs RGB normalisées sont généralement préférées [Crowley et Berard 1997 ; Graf et autres 1996 ; Hunke et Waibel 1994 ; Kawato et Ohya 2000 ; Koh et autres 1999 ; Ramos et autres 1997 ; Sun et autres 1998 ; Vincent et autres 1992 ; Wang et Sung 1999 ; Wang et autres 1995 ; Yachi et autres 2000 ; Yang et Waibel 1996], de sorte que l'effet de la luminance puisse être filtré. Les couleurs normalisées peuvent être dérivées des composantes RGB originales comme suit :

$$r = \frac{R}{R + G + B} \quad (2.2)$$

$$g = \frac{G}{R + G + B} \quad (2.3)$$

$$b = \frac{B}{R + G + B} \quad (2.4)$$

Des équations 2.2, 2.3 et 2.4 on voit que:  $r + g + b = 1$ . Les couleurs normalisées peuvent être effectivement représentées en utilisant seulement les valeurs  $r$  et  $g$  puisque la valeur  $b$  peut être obtenue par:  $b = 1 - r - g$ . Dans l'analyse de la couleur de la peau, un histogramme de couleur basé sur  $r$  et  $g$  montre que la couleur du visage occupe un petit cluster dans l'histogramme [Yang et Waibel 1996]. En comparant l'information de couleur d'un pixel avec les valeurs  $r$  et  $g$  du cluster du visage, la probabilité de l'appartenance du pixel à une région de peau humaine peut être déduite.

Sans compter le modèle de couleur RGB, il existe actuellement plusieurs modèles qui sont utilisés dans la recherche de la détection de visage. Dans [Lee et autres 1996], la représentation de couleurs *HSI* (*Hue, Saturation, and Brightness*) a montré des avantages par rapport à d'autres modèles en donnant une grande variance parmi les clusters de couleur des traits faciaux. Par conséquent, ce modèle est utilisé pour extraire les traits faciaux tels que les lèvres, les yeux et des sourcils. Puisque la représentation se relie fortement à la perception humaine de la couleur [Lee et autres

1996 ; Sobottka et Pitas 1996], elle est également largement répandue dans les algorithmes de segmentation de visage [Graf et autres 2000 ; Kapfer et Benoit-Pineau 1997 ; Luthon et Lievin 1997 ; McKenna et autres 1996 ; Shpungin et Movellan 2000 ; Sobottka et Pitas 1996 ; Terrillon et autres 1998 ; Woo et Oh 1999].

Le modèle de couleurs YIQ a été appliqué à la détection de visage dans [Dai et Nakano 1996 ; Wei et Sethi 1999]. En convertissant les couleurs RGB en représentation YIQ, on a constaté que la composante I qui inclut la couleur s'étendant de l'orange au cyan, parvient à gérer la région de la peau des Asiatiques [Dai et Nakano 1996]. La conversion supprime efficacement l'arrière plan des autres couleurs et permet la détection de petits visages dans un environnement naturel. D'autres modèles de couleur appliqués à la détection de visage incluent le modèle HSV [Garcia et Tziritas 1999 ; Herpers et autres 1999 ; Jordão et autres 1999 ; Yang 2000], le modèle YES [Saber et Tekalp 1998], le modèle YCrCb [Albiol 2000 ; Garcia et Tziritas 1999 ; Karlekar et Desai 1999 ; Menser et Brunig 1999 ; Tsapatsoulis et autres 2000 ; Wang et Chang 1997], le modèle YUV [Abdel-Mottaleb et Elgammal 1999 ; Marqués et Vilaplana 2000], le modèle CIE-xyz [Chen et Chiang 1997], le modèle  $L^*a^*b^*$  [Cai et Goshtasby 1999 ; Li et autres 2000], le modèle  $L^*u^*v^*$  [Hongo et autres 2000], le modèle CSN (une représentation modifiée de rg) [Kim et Kim 2000 ; Kim et autres 1998] et le modèle UCS/Farnsworth (un système de couleur perceptuellement uniforme proposé par Farnsworth [Wyszecki et Stiles 1967]) [Wu et autres 2000].

[Terrillon et autres 2000] ont récemment présenté une étude comparative de plusieurs espaces de couleurs largement utilisés (ou plus convenablement appelés les espaces de chrominance dans ce contexte puisque tous les espaces cherchent l'invariance de la luminance) pour la détection de visage. Dans leur étude, ils ont comparé l'espace TSL normalisé (*Teinte-Saturation-Luminance* [Terrillon et autres 2000]), les espaces de chrominance rg et CIE-xy, CIE-DSH, HSV, YIQ, YES, CIE- $L^*u^*v^*$  et CIE  $L^*a^*b^*$  en modélisant des distributions de la couleur de la peau avec un modèle Gaussien simple ou un modèle de mélange de densités Gaussiennes dans chaque espace. Les moments de *Hu* [Hu 1962] sont utilisés comme traits et un réseau de neurones multicouche à perceptron est appris pour classifier les candidats de visage. En général, ils ont prouvé que la couleur de la peau dans les espaces normalisés de chrominance peut être modélisée avec une simple Gaussienne et donner une très bonne performance, alors qu'un mélange de Gaussiennes est nécessaire pour les espaces non normalisés. Dans leur test de détection de visage, l'espace normalisé TSL fournit les meilleurs résultats, mais la conclusion générale est que le critère le plus important pour la détection de visage est le degré de chevauchement entre la peau et les distributions de non peau dans un espace donné (et ceci dépend fortement du nombre d'échantillons de peau et de non peau disponibles pour l'apprentissage).

La segmentation de couleur peut être fondamentalement effectuée en utilisant les seuils appropriés de la couleur de la peau où la couleur de la peau est modélisée par les histogrammes ou les diagrammes [Cai et Goshtasby 1999 ; Hongo et autres 2000 ; Kawato et Ohya 2000 ; Luthon et Lievin 1997 ; Sobottka et Pitas 1996 ; Wang et autres 1995 ; Woo et Oh 1999]. Des méthodes plus complexes se servent des mesures statistiques qui modélisent la variation de visage dans un spectre large [Albiol et autres 1999 ; Crowley et Berard 1997 ; Hunke et Waibel 1994 ; McKenna et autres 1996 ; Oliver et autres 2000 ; Yang et Waibel 1996]. Par exemple, [Oliver et autres 2000] et [Yang et Waibel 1996] utilisent une distribution Gaussienne pour

représenter un cluster de couleur de peau de milliers d'échantillons de couleurs de peau prélevés de races différentes. La distribution Gaussienne est caractérisée par sa moyenne ( $\mu$ ) et sa matrice de covariance ( $\Sigma$ ). La couleur d'un pixel d'une image d'entrée peut être comparée au modèle de couleur de peau en calculant la distance de Mahalanobis. Cette mesure de distance donne alors une idée de la façon dont la couleur du pixel ressemble à la couleur de peau du modèle.

Un avantage du modèle statistique de couleur est que la variation de couleur de nouveaux utilisateurs peut être adaptée dans le modèle général par une approche d'apprentissage. En utilisant une méthode adaptative, la détection de couleur peut être plus robuste par rapport à des changements de facteurs d'environnement tels que les conditions d'illumination et les caractéristiques de la caméra. Des exemples d'une approche d'apprentissage ont été utilisés par [Oliver et autres 2000] et par [Yang et Waibel 1996] en accord avec l'équation 2.6 qui met à jour les paramètres de la distribution Gaussienne [Oliver et autres 2000] (une approche similaire peut être trouvée dans le système de reconnaissance de visage de [McKenna et autres 1998]).

$$\Sigma_{new} = [\Sigma_{general}^{-1} + \Sigma_{user}^{-1}]^{-1} \quad (2.5)$$

$$\mu_{new} = \Sigma_{new} [\Sigma_{general}^{-1} \times \mu_{general} + \Sigma_{user}^{-1} \times \mu_{user}] \quad (2.6)$$

#### **2.4. L'approche basée sur l'image:**

On a montré dans la section précédente que la détection de visage par la modélisation explicite des traits faciaux est troublée par l'imprévisibilité de l'apparence du visage et des conditions environnementales. Bien que certaines des tentatives récentes basées sur les traits aient amélioré la capacité de faire face à l'imprévisibilité, la plupart d'entre elles sont encore limitées à des visages frontaux ou quasi-frontaux (ou sont incluses en tant qu'une des techniques dans un système combiné). Il y a toujours un besoin de techniques qui peuvent fonctionner dans des scénarios plus hostiles tels que la détection de visages multiples avec des arrières plans intensivement encombrés. Cette condition a inspiré un nouveau secteur de recherches dans lequel la détection du visage est traitée comme un problème de reconnaissance de formes. En formulant le problème en tant qu'un apprentissage pour reconnaître une forme de visage à partir d'exemples, l'application spécifique de la connaissance de visage est évitée. Ceci élimine le potentiel de modéliser l'erreur due à la connaissance incomplète ou imprécise du visage. L'approche de base en reconnaissant des formes de visage passe par l'intermédiaire d'un procédé d'apprentissage qui classe des exemples dans des classes de prototype de visage et de non visage. La comparaison entre ces classes et un tableau 2D d'intensités lumineuses (par conséquent le nom "basée sur l'image") extraites à partir d'une image d'entrée permet de décider de l'existence d'un visage. Les approches basées sur l'image les plus simples se fondent sur l'assortiment de modèles (template matching) [Holst 2000, Liu et Wang 2000], mais ces approches ne sont pas si performantes que les techniques plus complexes présentées dans les sections suivantes.

La plupart des approches basées sur l'image appliquent une technique à fenêtre de balayage (window scanning technique) pour détecter des visages. L'algorithme à fenêtre de balayage est essentiellement une recherche approfondie de l'image d'entrée pour des endroits possibles de visage à toutes les échelles, mais il y a des variétés d'implémentation de cet algorithme pour presque tous les systèmes basés sur l'image.

Typiquement, la taille de la fenêtre de balayage, le taux de sous échantillonnage, la taille d'étape et le nombre d'itérations changent selon la méthode proposée et le besoin d'un système efficace du point de vue calculs.

Dans les trois sections suivantes, les approches basées sur l'image sont divisées en méthodes des sous-espaces linéaires, en réseaux de neurones et en approches statistiques. Dans chaque section on donnera une présentation des caractéristiques de certaines des méthodes proposées, et dans la section 2.4.4, on essaiera de faire une évaluation comparative basée sur des résultats rapporté sur un ensemble de données commun.

### **2.4.1. Méthodes des sous espaces linéaires:**

Les images des visages humains se situent dans un sous-espace de l'espace global d'image. Pour représenter ce sous-espace, on peut utiliser des approches neurales (comme décrit dans la prochaine section), mais il existe également plusieurs méthodes plus étroitement liées à l'analyse statistique multi variable et qui peuvent être appliquées. Dans cette section on décrira et présentera des résultats de quelques unes de ces techniques, y compris l'analyse en composantes principales (PCA), l'analyse linéaire discriminante (LDA) et l'analyse des facteurs (Factor Analysis, FA).

Vers la fin des années 1980, [Sirovich et Kirby 1987] ont développé une technique qui utilise la PCA pour représenter efficacement les visages humains. Etant donné un ensemble d'images faciales différentes, la technique trouve d'abord les composantes principales de la distribution des visages, exprimées en termes de vecteurs propres (de la matrice de covariance de la distribution). Chaque visage individuel dans l'ensemble des visages peut alors être approximé par une combinaison linéaire des plus grands vecteurs propres, généralement désignés sous le nom des faces propres, en utilisant des poids appropriés.

[Turk et Pentland 1991] ont développé plus tard cette technique pour la reconnaissance de visage. Leur méthode exploite la nature distincte des poids des visages propres dans la représentation des visages individuels. Puisque la reconstruction de visage par ses composantes principales est une approximation, une erreur résiduelle est définie dans l'algorithme comme mesure préliminaire de "ressemblance au visage (faceness)". Cette erreur résiduelle qu'ils ont nommée la "distance par rapport à l'espaces des faces (Distance-From-Face-Space, DFFS)" donne une bonne indication de l'existence de visage par l'observation des minimums globaux dans la carte de distance.

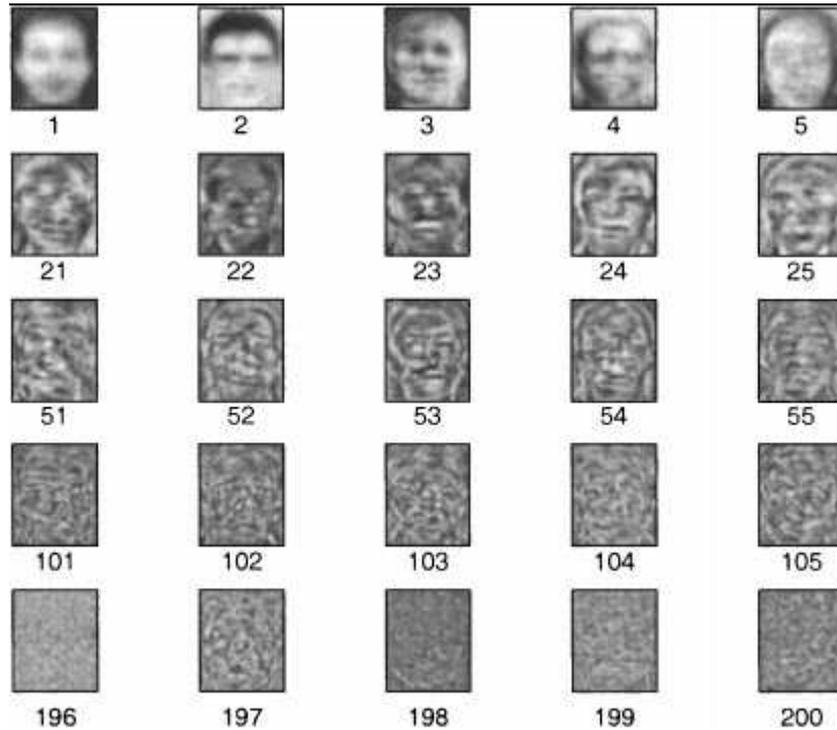
La procédure de base pour calculer l'espace des visages et la DFFS est comme suit [Turk et Pentland 1991]: On a un ensemble de données de  $n$  images  $\Gamma_1, \Gamma_2, \dots, \Gamma_n$ . Le visage moyen est défini par :

$$\Psi = \frac{1}{n} \sum_{i=1}^n \Gamma_i \quad (2.7)$$

Le visage moyen est ensuite soustrait de chaque image et chaque image est alors vectorisée :

$$\Phi_i = (\Gamma_i - \Psi)^v \quad (2.8)$$

Soit  $D = [\Phi_1 \Phi_2 \dots \Phi_n]$  et  $C = DD^t$ . Les vecteurs propres  $u_i$  de  $C$  sont appelés les composantes principales de  $D$ , lorsque on les reconvertit en matrices, on peut voir ces vecteurs comme des visages propres de l'ensemble de données. Quelques exemples de ces visages propres sont montrés en figure 2.4 (ordonnés selon les valeurs propres correspondantes). Ces vecteurs propres forment un sous-espace appelé espace des visages (face space).



**Figure 2.4** Quelques exemples des visages propres calculés à partir de la base de données ORL (le nombre sous chaque image indique le numéro d'ordre de la composante principale correspondante).

Une image d'entrée prétraitée  $\Phi$  peut être projetée dans l'espace des visages par :

$$\omega_k = u_k^t \Phi, \quad k = 1, \dots, m \quad (2.9)$$

Où  $m$  est le nombre de composantes principales choisies pour former l'espace des visages.  $m$  est inférieur à  $n$  puisque les composantes principales correspondantes aux valeurs propres les plus petites ne portent pas d'information significative pour la représentation (on considère par exemple le cas du visage propre numéro 200 dans la figure 2.4).

$\Phi$  peut être reconstruite par:

$$\Phi_r = \sum_{k=1}^m \omega_k u_k \quad (2.10)$$

L'erreur de reconstruction :  $\varepsilon = \|\Phi - \Phi_r\|^2$  est la DFFS.

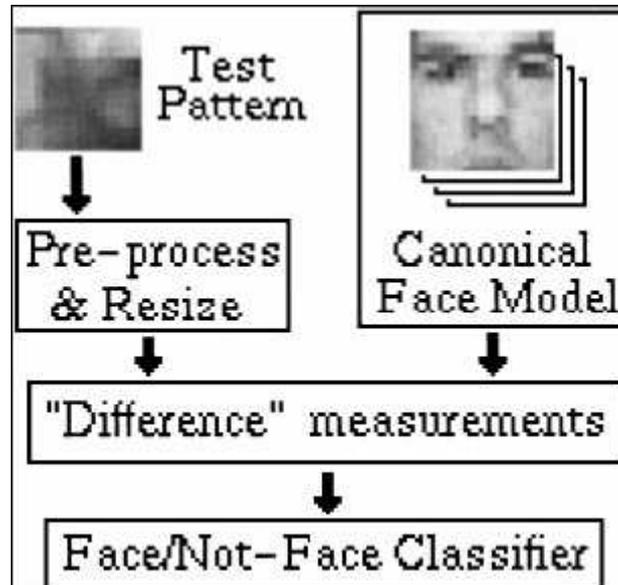
[Moghaddam et Pentland 1994 ; Pentland et autres 1994] ont proposé plus tard un détecteur de traits faciaux en utilisant la DFFS générée à partir des traits propres ou eigenfeatures (eigeneyes, eigennose, eigenmouth) obtenus à partir de divers modèles de traits faciaux dans un ensemble d'apprentissage. Le détecteur de traits a également une meilleure habilité à gérer les traits sous différents angles de vue puisque des traits de différentes vues discrètes ont été utilisés pendant l'apprentissage. La performance des positions des yeux est de 94% avec 6% de détections fausses positives [Pentland et autres 1994], dans une base de données de 7562 images de visages frontaux sur un arrière plan uni. 128 visages sont prélevés de la base de données pour calculer un ensemble de traits propres (eigenfeatures) correspondants. Une performance légèrement réduite mais encore précise pour les positions du nez et de la bouche a été également obtenue dans [Pentland et autres 1994]. La mesure de la DFFS a été également utilisée pour la détection des traits faciaux dans [Demirel et autres 1996] et en combinaison avec la discrimination linéaire de Fisher [Fisher 1936] pour la détection du visage et des traits faciaux [Senior 1999]. Une comparaison de la DFFS avec un algorithme d'empreintes digitales morphologique et multi échelles est fournie dans [Raducanu et Grana 2000].

Plus récemment, *Moghaddam* et *Pentland* ont encore développé cette technique dans un cadre probabiliste [Moghaddam et Pentland 1997]. Lors de l'utilisation de la PCA pour la représentation, on évite normalement le complément orthogonal de l'espace de visage (comme cité précédemment). *Moghaddam* et *Pentland* ont constaté que ceci mène à la supposition que l'espace de visage a une densité uniforme, ainsi ils ont développé un détecteur de maximum de vraisemblance qui tient compte de l'espace de visage et de son complément orthogonal pour gérer des densités arbitraires. Ils rapportent un taux de détection de 95% sur un ensemble de 7000 images de visage en détectant l'oeil gauche. Comparé au détecteur de DFFS, c'était sensiblement meilleur. Sur une tâche de détection multi échelles de têtes de 2000 images de visage de la base de données FERET, le taux de détection était de 97%. Dans [Jebara et Pentland 1997], *Jebara* et *Pentland* ont inclus cette technique dans un système pour dépister les visages humains, qui a été également basé sur la couleur, le 3D et l'information du mouvement. Une approche semblable est présentée dans [Kervrann et autres 1997] où la PCA est appliqué pour modéliser la classe des visages et la classe des pseudo visages (des formes de non visages, mais semblables aux visages), ainsi que des critères d'assortiment basés sur un rapport généralisé de probabilité. En [Meng et Nguyen 2000], la PCA est employée pour modéliser les visages et la complexité des arrières plans (un espace de visages propres (eigenface space) et un espace d'arrières plans complexes (eigenclutter space)).

[Samal et Iyengar 1995] ont proposé un schème PCA de détection basé sur des silhouettes de visage. Au lieu des faces propres, ils produisent des silhouettes propres (eigensilhouettes) et combinent ceci avec des techniques standard de traitement d'images (détection de bords, amincissement, seuillage,...) et la transformée généralisée de *Hough*. Ils rapportent un taux de détection de 92% sur un ensemble de données de 129 images (66 images de visages et 63 images générales) où, dans le cas des images de visage, les visages occupent la majeure partie de l'image.

La PCA est un moyen intuitif et approprié pour construire un sous-espace afin de représenter une classe d'objet dans plusieurs cas. Cependant, pour modéliser la variété des images des visages, la PCA n'est pas nécessairement optimale. L'espace

des visages pourrait mieux être représenté en le divisant en sous-classes, plusieurs méthodes ont été proposées pour ceci, dont la plupart sont basées sur un certain mélange de Gaussiennes multidimensionnelles. Cette technique a été appliquée pour la première fois pour la détection de visage par [Sung et Poggio 1998]. Leur méthode consiste principalement en quatre étapes (figure 2.5) :



**Figure 2.5** Classification dans le système de [Sung et Poggio 1998].

1. L'image d'entrée est prétraitée en réduisant sa taille à 19x19 pixels, en appliquant un masque pour éliminer les pixels proches des frontières, en soustrayant un plan d'intensité adapté des pixels de la fenêtre non masquée et en appliquant finalement l'égalisation de l'histogramme (figure 2.5).

2. Un modèle basé sur la distribution des formes canoniques des visages et des non visages est construit. Le modèle se compose de 12 clusters gaussiens multidimensionnels avec une position centroïde et une matrice de covariance, où 6 d'entre eux représentent des prototypes de formes de visage et 6 autres représentent des prototypes de formes de non visage. Les clusters sont construits par un algorithme de groupement elliptique des k-moyennes qui utilise une métrique de distance de Mahalanobis normalisée et qui change de manière adaptative.

3. Un ensemble de mesures d'image est calculé pour de nouvelles images relativement au modèle canonique de visage. Pour chacun des clusters, deux valeurs sont calculées. Une est la distance Mahalanobis entre la nouvelle image et le centroïde du prototype, définie dans le sous-espace formé par les 75 plus grands vecteurs propres du cluster du prototype, alors que l'autre est la distance Euclidienne de la nouvelle image à sa projection dans le sous-espace.

4. Un perceptron multicouche (MLP) est appris pour la classification de visage/non visage à partir du vecteur 24-dimensionnel des mesures de l'image. Le MLP n'est pas entièrement connecté, mais exploite quelques connaissances a priori du domaine.

L'ensemble d'apprentissage se compose de 47.316 vecteurs de mesures d'image, où 4150 sont des exemples de formes de visage.

Quand une nouvelle image est à classifier, les étapes de 1 jusqu'à 3 sont appliquées et le MLP donne la classification.

Une approche similaire à celle de *Sung* et *Poggio*, basée sur des traits des niveaux de gris, en combinaison avec des traits de texture a été exploré dans [Duta et Jain 1998], une autre approche similaire mais plus efficace (du point de vue calculs) a été proposée par [Fouad et autres 2000]. *Gu* et *Li* [Gu et Li 2000] ont présenté une autre variété de l'approche de *Sung* et *Poggio* où l'analyse linéaire discriminante est appliqué pour la sélection des traits avant l'apprentissage du classificateur à réseaux de neurones. Une autre méthode qui suit le cadre de travail de *Sung* et *Poggio* a été aussi proposée par [Rajagopalan et autres 1998] en utilisant des statistiques d'ordre élevé pour modéliser les clusters de visage et de non visage. Ils ont présenté également un nouvel algorithme de groupement en utilisant des statistiques d'ordre élevé qui remplacent l'algorithme de groupement des k-moyennes du travail de *Sung* et *Poggio*. Ils ont rapporté de bons résultats sur un sous-ensemble d'images de la base de données CMU en utilisant la technique des statistiques d'ordre élevé et en comparant ceci à un système basé sur les modèles cachés de Markov qui ne donne pas les mêmes performances. Les modèles cachés de Markov ont été également appliqués à la détection de visage dans [Marchand-Maillet et Merialdo 1999 ; Meng et autres 2000].

Une issue qui surgit lors de l'apprentissage des systèmes de reconnaissance de formes pour la classification visage/non visage est comment collecter un ensemble d'apprentissage qui sera représentable pour les images de la classe non visage. L'ensemble d'échantillons positifs d'apprentissage est facilement défini en tant que toutes sortes d'images de visage, mais il est plus difficile de définir l'ensemble complémentaire. *Sung* et *Poggio* ont suggéré un algorithme d'apprentissage, connu sous le nom de "apprentissage en circuit fermé (boot-strap training)" pour traiter partiellement ce problème (une stratégie plus précise que celle proposée dans [Burel et Carel 1994]). L'algorithme comprend les étapes suivantes :

1. Créer tout simplement l'ensemble initial des images de non visage en générant des images de pixels aléatoires,
2. Effectuer l'apprentissage du système,
3. Exécuter le système sur des scènes qui ne contiennent pas de visages et extraire les détections fausses positives,
4. Effectuer le prétraitement des détections fausses positives et leur ajout à l'ensemble d'apprentissage des non visages, aller à l'étape 2.

[Yang et autres. 2000] ont proposé deux méthodes pour la détection de visage qui cherchent également à représenter la diversité des visages humains comme ensemble de sous-classes. La première méthode est basée sur la FA, qui est une technique statistique multi variable tout à fait semblable à la PCA, mais contrairement à la PCA, la FA suppose que les échantillons de données observés viennent d'un modèle bien défini [Mardia et autres 1979].

$$x = \Lambda f + u + \mu \quad (2.11)$$

Où  $\Lambda$  est une matrice de constantes,  $f$  et  $u$  sont des vecteurs aléatoires et  $\mu$  est la moyenne. La FA cherche à trouver  $\Lambda$  et la matrice de covariance de  $u$  qui modélisent le mieux possible la structure de covariance de  $x$ . Si les variances spécifiques  $u$  sont supposées être nulles, la procédure de la FA peut être équivalente à la PCA. *Yang et autres* ont utilisé une mixture de la FA dans leur première méthode. Étant donné un ensemble d'images d'apprentissage, l'algorithme EM [Dempster et autres 1977] est utilisé pour estimer les paramètres dans le modèle de mixture de la FA [Ghahramani et Hinton 1996]. Ce modèle de mixture est ensuite appliqué à des sous-fenêtres dans l'image d'entrée et donne la probabilité qu'un visage soit présent dans la région en question.

Dans la deuxième méthode, *Yang et autres* ont utilisé la carte auto-organisatrice de Kohonen (Kohonen's Self-Organizing Map, SOM) [Kohonen 1995] pour diviser les images d'apprentissage en 25 classes de visage et 25 classes de non-visage (le nombre de classes est choisi selon la taille de l'ensemble d'apprentissage). Les images d'apprentissage sont ensuite utilisées pour calculer la LDA, qui n'est autre qu'une technique de réduction de dimension similaire à la PCA. Au lieu de calculer la matrice de transformation  $([u_1^t u_2^t \dots u_m^t])$  à partir de l'équation 2.9 pour la représentation de l'ensemble de données en entier (PCA), la LDA cherche à trouver une matrice de transformation qui est basée sur la maximisation de la dispersion interclasses et sur la minimisation de la dispersion intra-classes. Le vecteur propre dans la matrice de transformation de la LDA correspondant à la valeur propre la plus grande est connu sous le nom de *discriminant linéaire de Fisher (Fisher's linear discriminant)* [Fisher 1936], qui à son tour, a été utilisé en détection de visage [Song et Robinson 2000 ; Weber et Herrera Hernandez 1999]. La PCA a pour but la représentation tandis que la LDA a pour objectif la discrimination et est donc plus appropriée pour la détection de visage lorsque les classes de visage et de non-visage sont divisées en sous-classes. Avec cette méthode, les images d'entrée sont projetées dans un espace 49-dimensionnel  $(25+25-1)$ , dans lequel des distributions Gaussiennes sont utilisées pour modéliser la fonction de densité conditionnelle de chaque classe  $P(z / X_i)$ , où  $z$  est l'image projetée et  $i = 1, \dots, 49$ . Un visage est détecté ou n'est pas détecté dans une fenêtre de l'image d'entrée en se basant sur la règle de décision du maximum des probabilités :

$$X^* = \arg \max_{X_i} P(z / X_i) \quad (2.12)$$

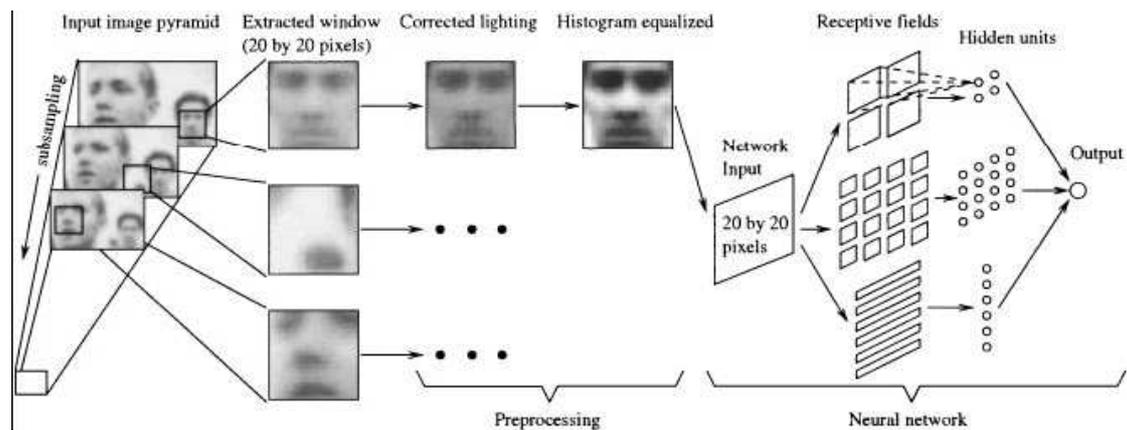
Chacune des deux méthodes de *Yang et autres* utilise une technique de fenêtre de 20 x 20 pixels, avec balayage de l'image d'entrée pendant 10 itérations et avec facteur de sous-échantillonnage de 1.2. D'autres approches de détection de visage qui utilisent la LDA sont présentées dans [Hotta et autres 1998] et [Tanaka et autres 1998], tandis que les SOMs ont été utilisées dans [Takacs et Wechsler 1997]. La règle de décision de probabilité a été récemment appliquée dans un système différent de détection de visage basé sur la décomposition des paquets d'ondelettes (wavelet packet decomposition) [Zhu et autres 2000].

### **2.4.2. Réseaux de neurones:**

Les réseaux de neurones sont devenus une technique populaire pour des problèmes de reconnaissance de formes, y compris la détection de visage. Les réseaux de neurones sont aujourd'hui beaucoup plus que de simples MLPs. Les architectures modulaires, la classification, les algorithmes d'apprentissage complexes, les réseaux de compression, les réseaux auto associatifs et les réseaux évolués ou combinés avec des algorithmes génétiques sont tous les exemples de l'utilisation répandue des réseaux de neurones dans la reconnaissance de formes. Pour la reconnaissance du visage, ceci implique que des approches neuronales pourraient être appliquées pour toutes les parties du système, et ceci avait été en effet montré dans plusieurs articles [Chellappa et autres 1995 ; Lin et autres 1997 ; Samal et Iyengar 1992]. Une introduction à quelques méthodes de base des réseaux de neurones pour la détection de visage peut être trouvée dans [Viennet et Fogelman Soulié 1998].

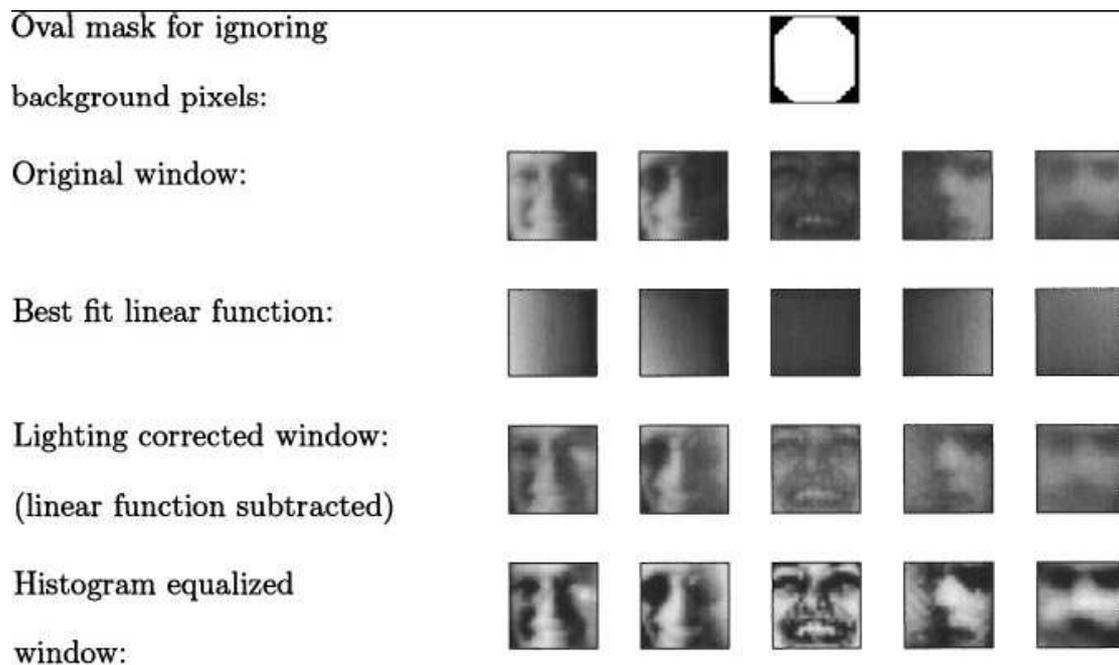
Les premières approches neuronales pour la détection de visage ont été basées sur les MLPs [Burel et Carel 1994 ; Juell et Marsh 1996 ; Propp et Samal 1992], où des résultats prometteurs ont été rapportés sur des ensembles de données assez simples. La première approche neuronale avancée qui a rapporté des résultats sur un ensemble de données grand et complexe (difficile) était par Rowley et autres. [Rowley et autres 1998]. Leur système incorpore la connaissance du visage dans un réseau de neurones connecté comme une rétine (retinally connected) comme montré dans la figure 2.6. Le réseau de neurones est conçu pour travailler avec des fenêtres de 20 x 20 pixels (ainsi 400 unités d'entrée). Il comprend une couche cachée avec 26 unités, où 4 unités travaillent avec des régions de 10 x 10 pixels, 16 autres travaillent avec des régions de 5 x 5 pixels et les 6 autres travaillent avec des fenêtres de 20 x 5 pixels recouvrant des raies horizontales. La fenêtre d'entrée est prétraitée par une correction de l'éclairage et par égalisation d'histogramme. Cette méthode de prétraitement a été adoptée dans le système de *Sung et Poggio* cité précédemment et est montrée dans la figure 2.7. Un problème qui surgit avec les techniques des fenêtres de balayage et le problème des détections recouvrantes. Rowley et autres ont traité ce problème par deux heuristiques :

1. Seuillage : le nombre de détections dans un petit voisinage entourant la l'endroit courant est compté, et s'il est au-dessus d'un certain seuil, un visage est présent à cet endroit.
2. Élimination du chevauchement : quand une région est classifiée comme visage selon le seuillage, alors les détections chevauchantes sont susceptibles d'être des détections fausses positives et sont ainsi rejetées.



**Figure 2.6** Le système de [Rowley et autres 1998].

Pour améliorer encore plus les performances, ils ont effectué l'apprentissage de multiples réseaux de neurones et ont combiné la sortie avec une stratégie d'arbitrage (ANDing, ORing, voting, ou un réseau de neurones d'arbitrage séparé). Cet algorithme a été appliqué dans un système de poursuite de personnes dans [Darrel et autres 1998] et [Sato 2000] et pour la détection initiale de visage dans le système de poursuite de tête de [La Cascia et autres 2000]. Un système similaire a été récemment proposé dans [Han et autres 2000].



**Figure 2.7** La méthode de prétraitement appliquée par [Rowley et autres 1998] et par Sung et Poggio. Une fonction linéaire est adaptée aux valeurs d'intensité dans la fenêtre (à l'intérieur du masque ovale) et ensuite soustraite (de la fenêtre entière). Ensuite l'égalisation de l'histogramme est appliquée pour améliorer le contraste.

Récemment, [Rowley et autres 1998b] ont combiné ce système avec un réseau de neurones (router neural network) pour détecter des visages à tous les angles dans le plan de l'image. Ils ont utilisé un MLP entièrement connecté avec une couche cachée et 36 unités de sortie (une unité pour chaque angle  $\pm 10$ ) pour décider de l'angle du visage. Le système a détecté 79,6% des visages dans deux grands ensembles de données avec un nombre restreint de détections positives fausses.

Dans [Feraud et autres 1997], Une approche neuronale différente a été suggérée, basée sur un modèle génératif avec contraintes (constrained generative model, CGM). Le CGM est un MLP auto associatif et entièrement connecté avec 3 couches de poids et avec 300 (15 x 20) unités d'entrée et de sortie (correspondant à la taille de l'image). La première couche cachée se compose de 35 unités tandis que la deuxième se compose de 50 unités. L'idée derrière ce modèle est de forcer l'exécution d'une PCA non linéaire en modifiant la projection des exemples de non visage de manière à être proches des exemples de visage. La classification est obtenue en considérant l'erreur de reconstruction du CGM (semblable à la PCA expliquée précédemment).

Pendant l'apprentissage, la cible pour une image de visage est la reconstruction de l'image elle-même, alors que pour des exemples de non visage, la cible est choisie comme la moyenne des  $n$  voisins les plus proches parmi les images de visages. *Feraud et autres* utilisent un algorithme d'apprentissage basé sur l'algorithme du circuit fermé de *Sung et Poggio* (et également une méthode de prétraitement semblable qui consiste en un lissage et une égalisation de l'histogramme). Pour mieux contrôler le processus d'apprentissage, ils utilisent une fonction de coût additionnelle basée sur la longueur de description minimale (Minimum Description Length, MDL). Le système est encore plus développé dans [Feraud et autres 2000] pour inclure l'information de la couleur et les vues multiples et est appliqué dans le problème de détection de visages sur le Web.

Dans [Lin et autres 1997], un système de reconnaissance de visage entièrement automatique est proposé, ce système est basé sur les réseaux de neurones basés sur une décision probabiliste (Probabilistic Decision-Based Neural Networks, PDBNN). Un PDBNN est un réseau de neurones de classification avec une structure modulaire hiérarchique. Le réseau est semblable au DBNN [Kung et Taur 1995], mais il a une contrainte probabiliste additionnelle. Le réseau se compose d'un sous réseau pour chaque classe d'objet, combiné avec une stratégie du type "le gagnant prend tout (winner-take-all)". Pour le cas de la détection de visage, le réseau se compose seulement d'un sous réseau représentant la classe de visage. L'apprentissage est effectué avec des règles d'apprentissage du DBNN, ce qui signifie que le professeur donne seulement l'exactitude de la classification (pas de valeurs de sortie exactes) et un apprentissage du type "globalement supervise, localement non supervise (Locally Unsupervised Globally Supervised, LUGS)" est appliqué. Avec le LUGS, l'apprentissage de chaque sous réseau est effectué individuellement avec un algorithme d'apprentissage non supervisé (Les k-moyennes et la quantification de vecteurs ou l'algorithme EM). L'apprentissage global est effectué pour affiner les frontières de décision par l'utilisation de l'apprentissage renforcé ou anti-renforcé lorsqu'une forme dans l'ensemble d'apprentissage est mal classifiée. Les images d'entrée sont à l'origine de taille 320 x 240 pixels (de l'ensemble de données MIT

[Sung et Poggio 1998]), mais sont réduites approximativement à une taille de 46 x 35 pixels, une fenêtre de taille 12 x 12 pixels est utilisée pour balayer cette image (avec une étape de recherche de 1 pixel).

Une nouvelle architecture apprentissage appelée SNoW (Sparse Network of Winnows) [Roth 1999] est appliquée à la détection de visage dans [Roth et autres 2000]. La SNoW, pour la détection de visage, est un réseau de neurones se composant de deux unités linéaires de seuil (Linear Threshold Units, LTU) (représentant les classes de visage et de non visage). Les deux LTUs opèrent dans un espace d'entrée de traits booléens. Le système le plus performant dérive des traits à partir de fenêtres de taille 20 x 20 pixels de la façon suivante : pour des fenêtres 1 x 1, 2 x 2, 4 x 4 et 10 x 10, calculer {position x intensité moyenne x variance de l'intensité}. Ceci donne des traits booléens dans un espace de traits de dimension 135,424, puisque la moyenne et la variance ont été discrétisées à un nombre prédéfini de classes. Les LTUs sont séparées l'une de l'autre et sont connectées d'une manière éparpillée dans l'espace des traits. Le système est appris avec une règle d'apprentissage simple qui favorise et défavorise les poids dans les cas de mauvaises classifications. D'une manière semblable aux méthodes précédemment mentionnées, *Roth et autres* utilisent la méthode du circuit fermé de *Sung et Poggio* pour produire des échantillons d'apprentissage et pré traitent toutes les images avec l'égalisation d'histogramme.

Outre la classification de visage, les réseaux de neurones ont été également appliqués pour la classification des traits faciaux [Duchnowski et autres 1995 ; McKenna et autres 1996 ; Reinders et autres 1996 ; Vincent et autres 1992], une méthode pour améliorer le temps de détection pour les MLPs est présentée dans [Ben-Yacoub et autres 1999].

### **2.4.3. Approches statistiques:**

Outre les méthodes linéaires des sous espaces et les réseaux de neurones, il existe plusieurs autres approches statistiques pour la détection de visage. Les systèmes basés sur la théorie de l'information, les SVM (Support Vector Machine) et la règle de décision de Bayes sont présentés dans cette section.

En se basant sur des premiers travaux de la détection de visage par maximum de ressemblance [Colmenarez et Huang 1996], Colmenarez et Huang [Colmenarez et Huang 1997] ont proposé un système basé sur l'information relative de Kullback (divergence de Kullback). Cette divergence est une mesure non négative de la différence entre deux fonctions de densité de probabilité  $P_{X^n}$  et  $M_{X^n}$  pour un processus aléatoire  $X^n$  :

$$H_{P\|M} = \sum_{X^n} P_{X^n} \ln \frac{P_{X^n}}{M_{X^n}} \quad (2.13)$$

Pendant l'apprentissage, pour chaque paire de pixels dans l'ensemble d'apprentissage, un histogramme est utilisé pour créer des fonctions de probabilité pour les classes de visage et de non visage. Puisque les valeurs des pixels dépendent fortement des valeurs de leurs voisins,  $X_n$  est traitée comme un processus Markovien du premier ordre et les valeurs des pixels dans les images de niveau de gris sont re-quantifiées à quatre niveaux. Colmenarez et Huang utilisent un grand ensemble

d'images de taille 11 x 11 pixels de visages et de non visages pour l'apprentissage, le procédé d'apprentissage résulte en un ensemble de LUTs (Look Up Table) avec des rapports de probabilité. Pour améliorer encore plus la performance et les exigences en matière de calculs, les paires de pixels qui contribuent faiblement à la divergence globale sont éliminées des LUTs et ne sont pas utilisées dans le système de détection de visage. Dans [Colmenarez et Huang 1998], Colmenarez et Huang ont encore amélioré cette technique en incluant l'erreur du circuit fermé (décrite dans la section 2.4.2). Dans [Colmenarez et autres 1999], la technique a été incorporé dans un système de poursuite de visage en temps réel. Un système similaire a été développé par Lew et Huijsmans [Lew et Huijsmans 1996].

Dans [Osuna et autres 1997], une SVM [Vapnik 1995] est appliquée à la détection de visage. Le système proposé suit le même cadre que celui développé par Sung et Poggio [Sung et Poggio 1998], décrit dans la section 2.4.1 (le balayage des images d'entrée avec une fenêtre de taille 19 x 19 pixels). Une SVM avec un polynôme du deuxième degré comme fonction noyau (kernel function) est apprise avec un algorithme de décomposition qui garantit l'optimalité globale. L'apprentissage est effectué avec l'algorithme du circuit fermé et les images sont prétraitées avec le procédé montré dans la figure 2.7. Kumar et Poggio [Kumar et Poggio 2000] ont récemment incorporé l'algorithme SVM de Osuna *et autres* dans un système pour la poursuite et l'analyse des visages en temps réel. Ils ont appliqué l'algorithme SVM sur des régions segmentées de peau dans les images d'entrée pour éviter le balayage exhaustif. Les SVMs ont été également utilisées pour la détection multi vues de visages en construisant des SVMs séparées pour différentes parties de la sphère de vue [Ng et Gong 1999]. Dans [Terrillon et autres 2000b], les SVMs ont amélioré les performances du détecteur de visage comparées à l'utilisation précédente du perceptron multi couches (MLP).

[Scneiderman et Kanade 1998 ; Schneiderman et Kanade 2000] ont décrits deux détecteurs de visage basés sur la règle de la décision de Bayes (présentée comme un rapport de probabilités dans l'équation 2.14).

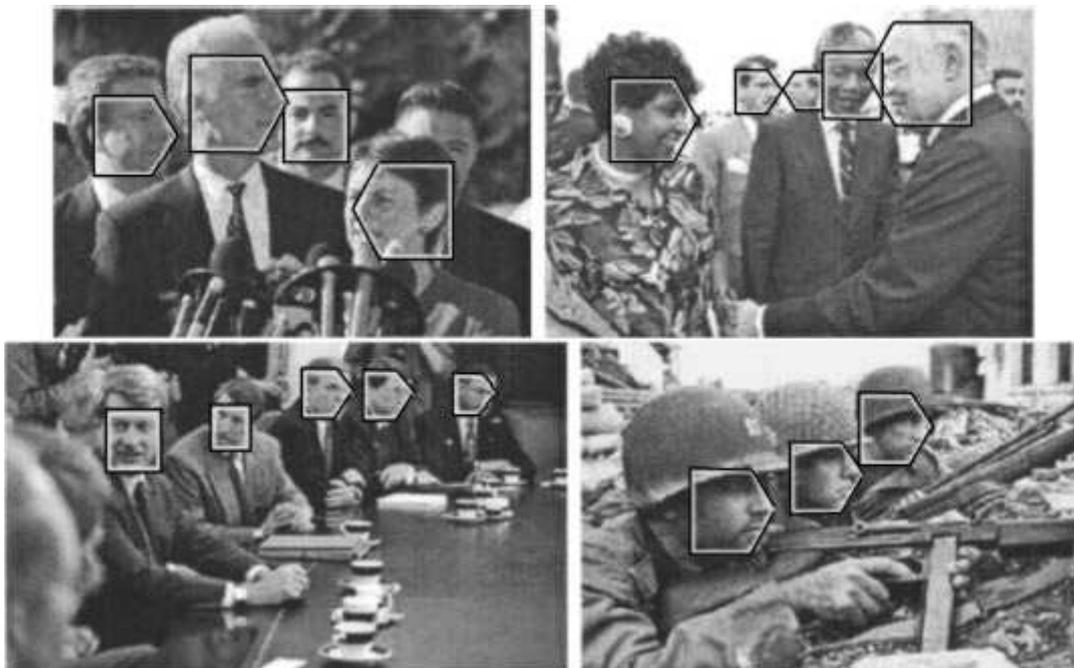
$$\frac{P(image | objet)}{P(image | non-objet)} > \frac{P(non-objet)}{P(objet)} \quad (2.14)$$

Si le rapport de probabilité (côté gauche) de l'équation 2.14 est plus grand que le côté droit, donc on décide qu'un objet (un visage) est présent à la position courante. L'avantage avec cette approche est que si les représentations pour  $P(image|objet)$  et de  $P(image|non-objet)$  sont précises, il s'avère que la règle de décision de Bayes est optimale [Duda et Hart 1973]. Dans le premier système de détection de visage proposé dans [Schneiderman et Kanade 1998], la fonction de probabilité à posteriori est dérivée en se basant sur un ensemble de modifications et de simplifications (dont certaines sont mentionnées ici) :

- La résolution de l'image de visage est normalisée à 64 x 64 pixels.
- Les images de visage sont décomposées en régions de 16 x 16 pixels et il n'existe aucune modélisation de la dépendance statistique entre les régions.
- Les régions sont projetées dans un sous espace de dimension 12 (construit par la PCA).

- La région entière du visage est normalisée pour avoir une moyenne nulle et une variance égale à l'unité.

Dans le deuxième système proposé en [Schneiderman et Kanade 2000], les attributs visuels de l'image ne sont pas représentés par des coefficients locaux de vecteurs propres (comme dans la première approche), mais par une transformée en ondelettes localement échantillonnée. La transformée en ondelettes peut prendre en charge l'information concernant les attributs visuels dans l'espace, la fréquence et l'orientation et devrait être plus appropriée pour décrire les caractéristiques du visage humain. La transformée en ondelettes appliquée dans [Schneiderman et Kanade 2000] est une décomposition à trois niveaux en utilisant un filtre linéaire de phase 5/3. Cette transformée décompose l'image en 10 sous bandes. De ces sous bandes, 17 attributs visuels (chacun consiste en 8 coefficients) sont extraits et traités en tant que variables aléatoires statistiquement indépendantes. Les coefficients sont re quantifiés à trois niveaux et les attributs visuels sont représentés en utilisant des histogrammes. Avec cette approche, un détecteur basé sur la vue est développé avec un détecteur de vue frontale et un détecteur de profil droit (pour détecter des images de profil gauche, le détecteur de profil droit est appliqué à des images renversées). Les meilleurs résultats parmi les deux systèmes décrits ici sont obtenus avec le système des vecteurs propres, mais ceci est dû à l'ensemble de données se composant majoritairement de visages de vue frontale. Dans une autre expérience menée sur un ensemble de données qui est composé dans sa majeure partie de visages de vue de profil, le détecteur à base d'ondelettes a surpassé le détecteur à vecteurs propres (qui avait été modifié pour détecter des vues de profil). Quelques exemples d'images traitées avec le système des ondelettes sont montrés dans la figure 2.8. La règle de décision de Bayes' a été également appliquée pour la détection de visage par [Qian et Huang 1997].



**Figure 2.8** Des exemples de détection de visage de [Schneiderman et Kanade 2000].

#### **2.4.4. Evaluation comparative:**

Puisque certains des systèmes basés sur l'image rapportent des résultats sur les mêmes ensembles de données, une comparaison de performances est donnée dans le tableau 2.1. Les ensembles de données ont été rassemblés à CMU par Rowley *et autres*. [Rowley et autres 1998] et au MIT par Sung et Poggio [Sung et Poggio 1998]. L'ensemble de données CMU inclut également l'ensemble de données MIT. Puisque quelques systèmes rapportent des résultats en excluant certaines des images des ensembles de données, chaque ensemble est considéré dans deux colonnes séparées dans le tableau 2.1. Ainsi, on traite les quatre ensembles de données suivants :

CMU-130: L'ensemble CMU total comprenant 130 images avec 507 visages étiquetés (qui inclut l'ensemble MIT). Les images sont à niveaux de gris et varient en qualité et en taille.

CMU-125: L'ensemble CMU en excluant les visages dessinés manuellement et ceux des cartoons pour un total de 125 images avec 483 visages étiquetés. Il existe aussi des visages additionnels dans cet ensemble de données (pour un total approximatif de 495 visages). Mais seulement 483 visages ont été retenus en se basant sur l'exclusion de certains des visages occlus et de visages non humains. Cependant, les 483 visages retenus par Rowley *et autres* ne sont pas nécessairement les mêmes 483 visages dans tous les résultats rapportés puisque dans (au moins) l'article [Roth et autres 2000], ils ont indiqué qu'ils ont eux-mêmes étiqueté l'ensemble des visages.

MIT-23 : L'ensemble de données entier de MIT (également connu sous le nom d'ensemble de test B de l'ensemble de données CMU) avec 23 images. Le nombre de visages étiquetés était à l'origine de 149 par Sung et Poggio (qui est le nombre utilisé dans les résultats rapportés dans [Sung et Poggio 1998] et [Lew et Huijsmans 1996]), mais il a été changé en 155 (qui est le nombre utilisé dans les résultats rapportés dans [Rowley et autres 1998], [Osuna et autres 1997] et [Lin et autres 1997]) une fois inclus dans l'ensemble de données CMU.

MIT-20: L'ensemble MIT en excluant les visages dessinés à la main et ceux des cartoons pour un total de 20 images avec 136 visages étiquetés.

**Tableau 2.1** Résultats rapportés en termes du pourcentage des détections correctes (CD) et du nombre de détections fausses positives (FP), CD/FP, sur les ensembles de données CMU et MIT ([Hjelmas et Low 2001]).

Système de détection	CMU-130	CMU-125	MIT-23	MIT-20
[Schneiderman et Kanade 2000]-E <sup>a</sup>		94.4%/65		
[Schneiderman et Kanade 2000]-W <sup>b</sup>		90.2%/110		
[Yang et autres 2000]-FA		92.3%/82		89.4%/3
[Yang et autres 2000]-LDA		93.6%/74		91.5%/1
[Roth et autres 2000]		94.8%/78		94.1%/3
[Rowley et autres 1998]	86.2%/23		84.5%/8	
[Feraud et autres 2000]	86%/8			
[Colmenarez et Huang 1997]	93.9%/8122			
[Sung et Poggio 1998]			79.9%/5	
[Lew et Huijsmans 1996]			94.1%/64	
[Osuna et autres 1997]			74.2%/20	
[Lin et autres 1997]			72.3%/6	
[Gu et Li 2000]			87.1%/0	

<sup>a</sup> coefficients des vecteurs propres ; <sup>b</sup> coefficients des ondelettes

Il est difficile de tirer des conclusions à partir du tableau 2.1, ceci est dû aux informations imprécises ou incomplètes dans les publications respectives. Il y a plusieurs problèmes qui doivent être abordés en détection de visages avec la technique de balayage par des fenêtres sur des données complexes telles que les ensembles de données CMU et MIT. Quelques uns de ces problèmes rencontrés sont les suivants:

Comment doit-on compter les détections correctes et les détections fausses positives ? Puisqu'une détection est un placement d'une fenêtre à un endroit dans l'image d'entrée, on doit décider à quelle précision doit répondre ce placement. Yang et autres. [Yang et autres 2000] ont pris cette décision en se basant sur la règle "...une détection de visage est une détection correcte si la région contient des yeux et une bouche." Pour tous les systèmes utilisant la technique de balayage par fenêtre, un visage pourrait être détecté à plusieurs échelles et à plusieurs positions proches l'une de l'autre. Rowley et autres et Gu et Li ont abordé ce problème en utilisant deux heuristiques de fusionnement (mentionnées dans section 2.4.2), alors que rares sont ceux qui semblent s'inquiéter à propos de ce sujet. C'est probablement dû au fait que les détections correctes et les détections fausses positives sont manuellement comptées pendant les tests.

Quelle est la courbe ROC du système ? Il est bien connu que quelques systèmes ont des taux élevés de détection tandis que d'autres ont un petit nombre de détections fausses positives. La plupart des systèmes peuvent ajuster leur seuil selon le degré de conservation du système voulu par l'utilisateur. Ceci peut être exprimé en termes de courbe ROC pour montrer le compromis entre détections correctes et détections fausses positives.

Quelle est la taille de l'ensemble d'apprentissage et comment l'apprentissage doit-il être implémenté ? Quelques systèmes utilisent un grand ensemble d'apprentissage et génèrent des échantillons additionnels d'apprentissage en tournant, en reflétant et en ajoutant du bruit, alors que d'autres ont un ensemble d'apprentissage

plus petit. En outre, l'algorithme d'apprentissage du circuit fermé proposé par Sung et Poggio est implémenté dans certains systèmes.

Qu'est ce que c'est qu'un visage ? Puisque l'ensemble de données CMU contient un grand nombre de visages, il semble y avoir un certain désaccord sur le nombre de visages que l'ensemble de données contient réellement. Ceci est dû au fait que des visages humains, des visages d'animaux, des visages de cartoons et des visages dessinés à la main sont présents dans l'ensemble de données.

Puisque plusieurs de ces questions sont laissées sans réponse dans la plupart des articles, ceci introduit un degré d'incertitude dans les performances des systèmes. Sans courbes ROC, il est difficile de dire à quel point les systèmes sont affectés par des ajustements de paramètres. En outre, le nombre de détections fausses positives dépend du nombre de fenêtres examinées avec la technique de balayage par fenêtres, ce qui rend dure l'évaluation du taux de détections fausses.

Cependant, tous les résultats mentionnés dans le tableau 2.1 sont assez impressionnants concernant la complexité des bases de données MIT et CMU.

### **2.5. Conclusion:**

Dans ce chapitre, on a présenté un historique des travaux effectués dans le domaine de la détection de visage, ensuite, on a passé en revue les étapes d'évolution de ce domaine, on a fourni une catégorisation des techniques utilisées pour la détection, la première catégorie englobe les approches dites "basées sur les traits" telles que l'analyse de bas niveau, la deuxième catégorie rassemble les approches dites "basées sur l'image" comme les méthodes des sous espaces linéaires, les réseaux de neurones,...etc. Enfin, on a présenté une évaluation comparative des systèmes basés sur les mêmes ensembles de données.

# Chapitre 3:

# La reconnaissance de visage



### **3.1. Introduction:**

En tant qu'une des applications les plus réussies de l'analyse et de la compréhension d'image, la reconnaissance de visage a récemment suscité une grande attention, particulièrement au cours des dernières années. Au moins deux raisons expliquent cette tendance : la première est l'éventail d'applications de commerce et de renforcement de la loi et la seconde est la disponibilité des technologies faisables après 30 ans de recherche. De plus, le problème de la reconnaissance du visage humain continue à attirer des chercheurs d'autres disciplines comme le traitement d'image, la reconnaissance de formes, les réseaux de neurones, la vision par ordinateur, la conception graphique par ordinateur et la psychologie.

Quoique les systèmes courants de reconnaissance aient atteint un certain niveau de maturité, leur succès est limité par les conditions imposées par les applications réelles. Par exemple, l'identification des images de visage acquises dans un environnement extérieur avec des changements de l'illumination et/ou de la pose demeure un problème en grande partie non résolu. En d'autres termes, les systèmes courants sont encore très loin des possibilités du système de perception humaine.

Le besoin grandissant de systèmes faciles à utiliser qui peuvent sécuriser nos capitaux et protéger notre intimité sans perdre notre identité dans une grande foule de nombres est évident. Actuellement, on a besoin d'un code PIN pour obtenir de l'argent de la banque, d'un mot de passe pour un ordinateur, d'une autre douzaine de mots de passe pour accéder à l'Internet et ainsi de suite. Bien que des méthodes très fiables d'identification biométrique d'individus existent -par exemple, l'analyse de l'empreinte digitale, de l'iris ou de la rétine-, ces méthodes se fondent sur la coopération des participants, tandis qu'un système d'identification de personnes basé sur l'analyse des vues frontales ou de profil du visage est souvent efficace sans même nécessiter la coopération ou la connaissance du participant. Une partie des avantages et des inconvénients de la biométrie peut être trouvé en [Phillips et autres 1998]. Le Tableau 3.1 énumère certaines des applications de la reconnaissance de visage.

Les applications de la reconnaissance de visage, commerciales ou de renforcement de la loi, s'étendent des photographies statiques à format contrôlé aux images vidéo non contrôlées, posant un éventail de défis techniques et exigeant également un éventail de techniques de traitement d'image, d'analyse, de compréhension, et de reconnaissance de formes. On peut généralement classifier les systèmes de reconnaissance de visage dans deux groupes selon qu'ils se servent d'images statiques ou de vidéos. A l'intérieur de chacun de ces groupes, des différences significatives existent, selon l'application mise en oeuvre. Les différences sont en termes de qualité d'image, de degré de complexité de l'arrière plan (qui pose de grands défis aux algorithmes de segmentation), de variations des images d'un individu particulier qui doit être identifié, de disponibilité d'un critère bien défini pour la reconnaissance ou l'identification, de la nature, du type et de la quantité d'information fournie par utilisateur. Une liste de quelques systèmes commerciaux est donnée dans le tableau 3.2.

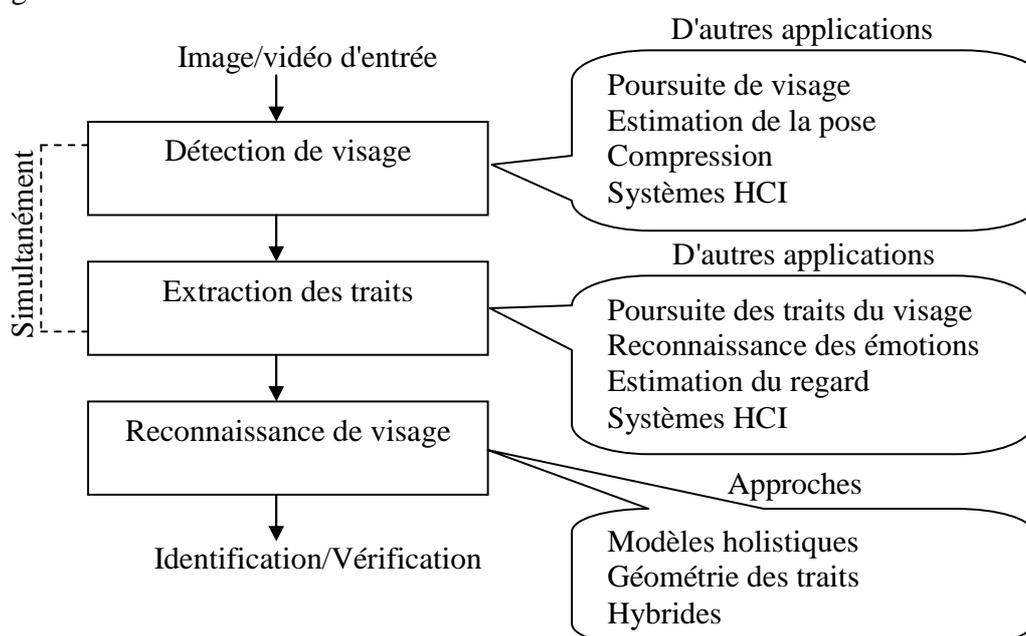
**Tableau 3.1** Applications typiques de la reconnaissance de visage.

<b>Domaines</b>	<b>Applications spécifiques</b>
Divertissement	Jeux vidéo, réalité virtuelle, programmes de formation
	Interaction homme-robot, Interaction homme-ordinateur
Cartes intelligentes	Permis de conduite,
	Immigration, ID nationaux, passeports, enregistrement d'électeurs
	Fraude d'assistance sociale
Sécurité de l'information	Contrôle parental de TV, ouverture personnelle de dispositifs ou de bureaux
	Sécurité des applications, Sécurité des bases de données, Encryptage de fichiers
	Sécurité Intranet, Accès Internet, Enregistrements médicaux
	Sécurité des terminaux publics
Renforcement de la loi et surveillance	Vidéo surveillance avancée
	Commande de portique, Analyse après événements
	Recherche et poursuite de suspects

**Tableau 3.2** Systèmes de reconnaissance de visage disponibles (Quelques sites Web ont peut être changé).

<b>Produits commerciaux</b>	<b>Sites Web</b>
FaceIt from Visionics	<a href="http://www.FaceIt.com">http://www.FaceIt.com</a>
Viisage Technology	<a href="http://www.viisage.com">http://www.viisage.com</a>
FaceVACS from Plettac	<a href="http://www.plettac-electronics.com">http://www.plettac-electronics.com</a>
FaceKey Corp.	<a href="http://www.facekey.com">http://www.facekey.com</a>
Cognitec Systems	<a href="http://www.cognitec-systems.de">http://www.cognitec-systems.de</a>
Keyware Technologies	<a href="http://www.keywareusa.com/">http://www.keywareusa.com/</a>
Passfaces from ID-arts	<a href="http://www.id-arts.com/">http://www.id-arts.com/</a>
ImageWare Software	<a href="http://www.iwsinc.com/">http://www.iwsinc.com/</a>
Eyematic Interfaces Inc.	<a href="http://www.eyematic.com/">http://www.eyematic.com/</a>
BioID sensor fusion	<a href="http://www.bioid.com">http://www.bioid.com</a>
Visionsphere Technologies	<a href="http://www.visionspheretech.com/menu.htm">http://www.visionspheretech.com/menu.htm</a>
Biometric Systems, Inc.	<a href="http://www.biometrica.com/">http://www.biometrica.com/</a>
FaceSnap Recoder	<a href="http://www.facesnap.de/htdocs/english/index2.html">http://www.facesnap.de/htdocs/english/index2.html</a>
SpotIt for face composite	<a href="http://spotit.itc.it/SpotIt.html">http://spotit.itc.it/SpotIt.html</a>

Un énoncé général du problème de la reconnaissance de visages peut être formulé comme suit : étant donné une image ou une vidéo d'une scène, identifier ou vérifier une ou plusieurs personnes dans la scène en utilisant une base de données de visages stockés. L'information collatérale disponible telle que la race, l'âge, le genre, le sexe, l'expression faciale ou la parole peut être utilisée pour rétrécir l'étendu de la recherche (amélioration de la reconnaissance). La solution à ce problème implique la segmentation des visages (détection de visage) à partir de scènes complexes, l'extraction des traits des régions du visage, la reconnaissance ou la vérification (Figure 3.1). Dans les problèmes de reconnaissance, l'entrée du système est un visage inconnu, et le système retourne son identité déterminée à partir d'une base de données d'individus connus, tandis que dans les problèmes de vérification, le système doit confirmer ou rejeter l'identité réclamée du visage d'entrée.



**Figure 3.1** Configuration d'un système de reconnaissance de visages.

La perception du visage est une partie importante des possibilités du système de perception humaine et est une tâche courante pour les humains, alors qu'établir un système informatique semblable est toujours un secteur en cours de recherches. Les premiers travaux sur la reconnaissance de visage datent au moins des années 50 en psychologie [Bruner et Tagiuri 1954] et aux années 60 dans la littérature de l'ingénierie [Bledsoe 1964]. Certaines des études les plus anciennes incluent un travail sur l'expression faciale des émotions par [Darwin 1972] (voir également [Ekman 1998]) et sur la biométrie faciale basée sur le profil par [Galton 1888]. Mais la recherche sur la reconnaissance automatique de visage par machine n'a commencé vraiment que dans les années 70 [Kelly 1970] et après le travail séminal de [Kanade 1973]. Au cours des 30 dernières années, une recherche étendue a été conduite par des psycho physiciens, des neurologistes et des ingénieurs sur des aspects divers de la reconnaissance de visage par les hommes et par les machines. Psycho physiciens et neurologistes ont été concernés par

des issues comme: "est-ce que la perception du visage est un processus dédié (cette issue est discutée toujours dans la communauté des psychologues [Biederman et Kalocsai 1998 ; Ellis 1986 ; Gauthier et autres 1999 ; Gauthier et Logothetis 2000]) et est-ce qu'elle est accomplie d'une manière holistique ou par analyse de traits (caractéristiques) locaux".

Plusieurs des hypothèses et des théories proposées par des chercheurs dans ces disciplines ont été basées sur des ensembles d'images plutôt petits. Néanmoins, plusieurs des résultats obtenus ont eu des conséquences importantes pour les ingénieurs qui conçoivent des algorithmes et des systèmes pour la reconnaissance de visage par des machines.

Sauf quelques exceptions qui emploient des données de gamme [Gordon 1991], le problème de la reconnaissance de visage a été formulé en tant que problème de reconnaissance d'objets tridimensionnels (3D) à partir d'images bidimensionnelles (2D) (il y a eu des avancées récentes dans la reconnaissance tridimensionnelle (3D) du visage dans des situations où les données de gamme acquises à travers une lumière structurée peuvent être assorties sûrement [Bronstein et autres 2003]). Les premières approches l'ont traité comme un problème de reconnaissance de formes bidimensionnelles. En conséquence, pendant le début et le milieu des années 70, des techniques typiques de classification de formes, qui emploient des attributs mesurés de traits (par exemple, les distances entre les points importants) dans les visages ou les profils de visage, ont été employées [Bledsoe 1964 ; Kanade 1973 ; Kelly 1970]. Pendant les années 80, le travail sur la reconnaissance de visage est demeuré en grande partie dormant. Depuis le début des années 90, l'intérêt de recherches pour la technologie de reconnaissance de visage s'est développé considérablement. On peut attribuer ceci à plusieurs facteurs: une augmentation de l'intérêt pour les applications commerciales, la disponibilité de matériel "temps réel" et l'importance croissante des applications de surveillance.

Au cours des 15 dernières années, la recherche s'est concentrée sur la façon de rendre les systèmes de reconnaissance de visage entièrement automatiques en abordant des problèmes tels que la localisation d'un visage dans une image ou dans une vidéo données et l'extraction des traits, tels que les yeux, la bouche, etc.... Entre-temps, des avancées significatives ont été faites dans la conception des classificateurs pour une reconnaissance de visage réussie. Parmi les approches holistiques basées sur l'apparence, *les visages (faces) propres (eigenfaces)* [Kirby et Sirovich 1990 ; Turc et Pentland 1991] et *les visages (faces) de Fisher (Fisherfaces)* [Belhumeur et autres 1997 ; Etemad et Chellappa 1997 ; Zhao et autres 1998] se sont avérées efficaces dans les expériences avec de grandes bases de données. *Les approches d'assortiment de graphes basées sur les traits (feature-based graph matching methods)* [Wiskott et autres 1997] ont été également tout à fait réussies. Comparées aux approches holistiques, les méthodes basées sur les traits sont moins sensibles aux variations de l'illumination et du point de vue et à l'inexactitude dans la localisation du visage. Cependant, les techniques d'extraction de traits requises pour ce type d'approche ne sont toujours pas fiables ou assez précises [Cox et autres 1996]. Par exemple, la plupart des techniques de localisation des yeux assument quelques modèles géométriques et des textures qui ne fonctionnent pas si les yeux sont

fermés. La section 3.2 présentera un examen de la reconnaissance de visage basée sur des images statiques.

Pendant les 5 à 8 dernières années, la grande partie des recherches a été concentrée sur la reconnaissance de visage basée sur la vidéo. Le problème des images statiques a plusieurs avantages et inconvénients inhérents. Pour des applications telles que les permis de conduite, en raison de la nature contrôlée du procédé d'acquisition d'image, le problème de segmentation est plutôt facile. Cependant, si seulement une image statique d'une scène d'aéroport est disponible, la localisation et la segmentation automatiques d'un visage pourraient poser des défis sérieux à n'importe quel algorithme de segmentation. D'autre part, si une séquence vidéo est disponible, la segmentation d'une personne mobile peut être accomplie plus facilement en utilisant le mouvement comme information. Mais la taille réduite et la mauvaise qualité des images des visages capturés de la vidéo peuvent augmenter de manière significative la difficulté de la reconnaissance.

La reconnaissance d'un objet 3D à partir de ses images 2D pose beaucoup de défis. Les problèmes d'illumination et de pose sont deux issues importantes pour les approches basées sur l'apparence et pour celles basées sur l'image. Beaucoup d'approches ont été proposées pour manipuler ces issues, avec la plupart d'entre elles explorant la connaissance disponible du domaine.

En 1995, un article [Chellappa et autres. 1995] a donné un aperçu complet des technologies de reconnaissance de visage à ce moment-là. Un autre aperçu [Samal et Iyengar 1992] est apparu en 1992. A cette époque, la reconnaissance de visage basée sur la vidéo était toujours à une étape naissante. Pendant les 8 dernières années, la reconnaissance de visage a suscité une attention accrue et a beaucoup avancé. Beaucoup de systèmes commerciaux pour la reconnaissance de visage à partir d'images statiques sont maintenant disponibles. Récemment, des efforts significatifs de recherches ont été concentrés sur la modélisation, la poursuite et la reconnaissance du visage et à leur intégration dans des systèmes. De nouveaux ensembles de données (bases de données de visages) ont été créés et des évaluations des techniques de reconnaissance employant ces bases de données ont été proposées. Ce n'est pas une exagération de dire que la reconnaissance du visage est devenue l'une des applications les plus actives de la reconnaissance de formes, de l'analyse et de la compréhension d'images.

### **3.2. Reconnaissance de visage à partir d'images statiques:**

Comme illustré sur la figure 3.1, le problème de la reconnaissance automatique de visage implique trois étapes principales : (1) détection et normalisation approximative des visages, (2) l'extraction des traits et la normalisation précise des visages, (3) l'identification et/ou la vérification. Parfois, différentes tâches secondaires ne sont pas totalement séparées. Par exemple, les traits faciaux (yeux, nez, bouche) utilisés pour la reconnaissance du visage sont souvent employés dans la détection du visage. La détection du visage et l'extraction des traits peuvent être accomplies simultanément, comme indiqué sur la figure 3.1. Selon la nature de l'application, par exemple, les tailles des *ensembles d'apprentissage et de test (training and test sets)*, la complexité

(encombrement) et la variabilité de l'arrière plan, le bruit, l'occlusion et les exigences de vitesse d'exécution, certaines des tâches secondaires peuvent causer d'énormes défis.

Bien que les systèmes entièrement automatiques de reconnaissance de visage doivent exécuter chacune des trois tâches secondaires, la recherche sur chaque tâche secondaire est critique. Ceci est vrai non seulement parce que les techniques utilisées pour les différentes tâches secondaires doivent être améliorées, mais également parce qu'elles sont critiques dans différentes applications (figure 3.1). Par exemple, la détection de visage est nécessaire pour initialiser la poursuite du visage, l'extraction des traits faciaux est nécessaire pour identifier l'émotion humaine, qui est à son tour essentielle dans les systèmes de *l'interaction homme-ordinateur (HCI for Human Computer Interaction)*. L'isolement des tâches secondaires facilite l'évaluation et l'avancement de l'état de l'art des techniques composantes. Les premières techniques de détection de visage pouvaient seulement manipuler un seul visage à la fois ou un petit nombre de visages bien séparés et de vue frontale, dans des images ayant des arrière plans relativement simples, alors que les algorithmes de dernier cri peuvent détecter les visages et leurs poses dans des milieux encombrés [Gu et autres 2001 ; Heisele et autres. 2001 ; Schneiderman et Kanade 2000a ; Viola et Jones 2001]. Une recherche extensive sur les tâches secondaires a été effectuée et les aperçus appropriés sont apparus, par exemple, la tâche secondaire de la détection de visage [Hjelmas et al 2001 ; Yang et autres 2002].

### **3.2.1. Les étapes clés avant la reconnaissance du visage: La détection du visage et l'extraction des traits:**

La première étape dans tous les systèmes automatiques de reconnaissance de visage est la détection des visages dans les images. Après qu'un visage ait été détecté, la tâche de l'extraction des traits est effectuée dans le but d'obtenir les traits qui sont introduits dans un système de classification de visage. Selon le type du système de classification, les traits peuvent être des traits locaux tels que des lignes ou des points, ou des traits faciaux tels que les yeux, le nez et la bouche. La détection du visage peut également utiliser des traits, dans ce cas, ces traits sont extraits simultanément avec la détection du visage. L'extraction des traits est également une clé pour l'animation et la reconnaissance des expressions faciales.

Sans considérer les positions des traits, la détection du visage est jugée réussie si la présence et l'endroit approximatif d'un visage ont été correctement identifiés. Cependant, sans endroit précis du visage et des traits, on observe une dégradation apparente dans les performances de la reconnaissance [Martinez 2002 ; Zhao 1999]. La relation étroite entre l'extraction des traits et la reconnaissance du visage nous motive pour passer en revue quelques unes des méthodes d'extraction de traits qui sont employées dans les approches de reconnaissance et qui seront examinées dans la section 3.2.2. Par conséquent, cette section sert également d'introduction à la prochaine section.

#### **3.2.1.1. Segmentation/Détection:**

Jusqu'au milieu des années 90, la plupart du travail sur la segmentation a été concentré sur la segmentation d'un visage unique à partir d'un arrière plan simple ou

complexe. Ces approches utilisaient un modèle du visage entier, un modèle déformable basé sur les traits, la couleur de la peau et un réseau de neurones.

Des avancées significatives ont été faites ces dernières années en réalisant la détection automatique du visage dans diverses conditions. Comparées aux méthodes basées sur les traits et aux méthodes d'assortiment de modèles (*template-matching methods*), les méthodes basées sur l'apparence ou l'image [Rowley et autres 1998 ; Sung et Poggio 1997] qui se chargent de l'apprentissage des systèmes sur de grands nombres d'échantillons ont donné les meilleurs résultats. Ceci peut ne pas être surprenant puisque les formes du visage sont compliquées, très semblables entre eux et différents des formes appartenant à ce que l'on puisse appeler "des non visages". A travers un apprentissage extensif, les ordinateurs peuvent être tout à fait bons pour détecter des visages.

Plus récemment, la détection des visages sous une rotation en profondeur (en dehors du plan de l'image bidimensionnelle) a été étudiée. Une approche est basée sur l'apprentissage sur des échantillons contenant des vues multiples [Gu et autres 2001 ; Schneiderman et Kanade 2000a]. Comparées aux méthodes basées sur des traits invariants [Wiskott et autres 1997], les méthodes de détection et de reconnaissance de visage basées sur des vues multiples semblent pouvoir réaliser de meilleurs résultats quand l'angle de la rotation en dehors du plan de l'image est grand (35°). Dans la communauté de psychologie, une discussion semblable existe sur le fait que l'identification de visage soit invariante par rapport au point de vue ou non. Les études dans les deux disciplines semblent soutenir l'idée qui, pour de petits angles, dit que la perception de visage est indépendante de la vue, alors que pour de grands angles, elle est dépendante de la vue.

Dans un problème de détection, deux statistiques sont importantes : *les détections vraies positives* (également désignés sous le nom du *taux de détection*) et *les détections fausses positives* (les détections dans des régions de non visage). Un système idéal aurait un taux de détections vraies positives très élevé et un taux de détections fausses positives très bas. Dans la pratique, ces deux conditions sont en conflit. Le traitement de la détection de visage comme un problème de classification à deux classes aide à réduire dramatiquement le taux des détections fausses positives [Rowley et autres 1998 ; Sung et Poggio 1997] tout en maintenant le taux des détections vraies positives. Ceci est réalisé par un réapprentissage des systèmes avec les échantillons qui ont généré des détections fausses positives au cours des apprentissages précédents.

### **3.2.1.2. Extraction des traits:**

L'importance des traits faciaux pour la reconnaissance du visage ne peut pas être exagérée. Beaucoup de systèmes de reconnaissance de visage ont besoin des traits faciaux en plus du visage holistique, comme suggéré par des études en psychologie. Il est bien connu que même les méthodes holistiques d'assortiment, par exemple *eigenfaces* [Turk et Pentland 1991] et *Fisherfaces* [Belhumeur et autres 1997], ont besoin des positions précises des traits faciaux principaux tels que les yeux, le nez et la bouche pour normaliser le visage détecté [Martinez 2002 ; Yang et autres 2002].

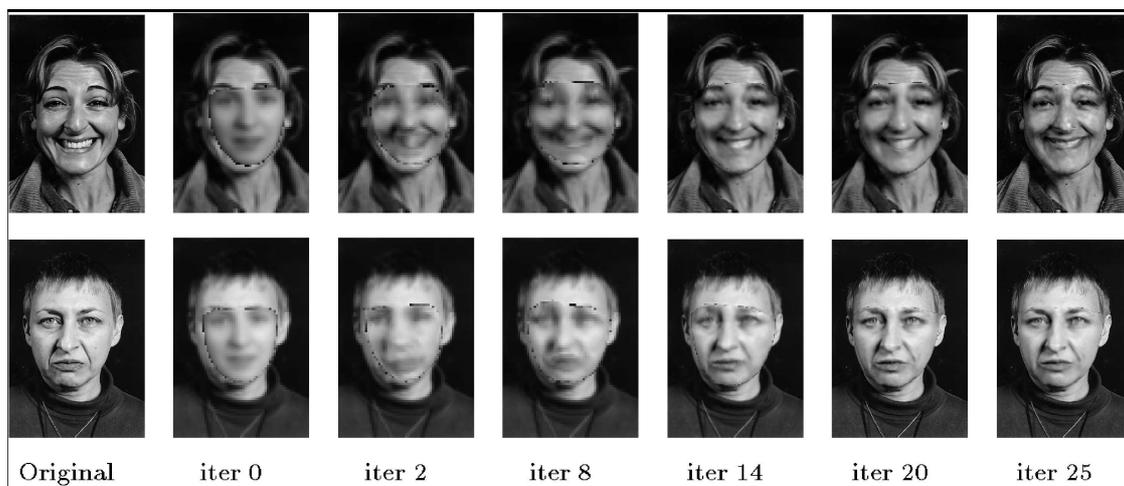
Trois types de méthodes d'extraction de traits peuvent être distingués: (1) les méthodes génériques basées sur les bords, les lignes et des courbes; (2) les méthodes basées sur les modèles de traits (*feature-template-based-methods*) qui sont employées pour détecter les traits faciaux tels que les yeux; (3) les méthodes structurelles d'assortiment (*structural matching methods*) qui prennent en considération les contraintes géométriques sur les traits. Les nouvelles approches sont concentrées sur des traits individuels; par exemple, une approche basée sur les modèles a été décrite dans [Hallinan 1991] pour détecter et identifier l'oeil humain dans un visage frontal. Ces méthodes ont des difficultés quand les apparences et les traits changent de manière significative, par exemple, les yeux fermés, les yeux avec des verres ou la bouche ouverte. Pour détecter les traits plus sûrement, les approches récentes ont employé des méthodes structurelles d'assortiment (*structural matching methods*), par exemple, *le modèle de forme active (Active Shape Model)* [Cootes et autres 1995]. Comparées aux anciennes méthodes, ces méthodes statistiques récentes sont beaucoup plus robustes en termes de gestion des variations de l'intensité de l'image et de la forme des traits.

Une situation bien plus défiante pour l'extraction des traits est la "restauration" de traits, qui essaye de récupérer les traits qui sont invisibles à cause des grandes variations de pose de la tête. Ici, la meilleure solution pourrait être de produire des hallucinations pour les traits manquants en employant la symétrie bilatérale du visage ou en employant l'information apprise. Par exemple, *une méthode statistique basée sur des vues multiples (view-based statistical method)* prétend même pouvoir manipuler les vues de profil dans lesquelles beaucoup de traits locaux sont invisibles [Cootes et autres 2000].

Une approche basée sur les modèles (*template-based approach*) pour détecter les yeux et la bouche dans des images de scènes réelles a été présentée dans [Yuille et autres 1992]. Cette méthode est basée sur l'assortiment d'un modèle paramétré et prédéfini à une image qui contient une région de visage. Deux modèles sont employés pour assortir les yeux et la bouche respectivement. Une fonction d'énergie est définie, elle lie les bords, les crêtes et les vallées dans l'image aux propriétés correspondantes dans le modèle, cette fonction d'énergie est réduite au minimum en changeant itérativement les paramètres du modèle pour s'adapter à l'image. Comparé à ce modèle, qui est conçu manuellement, le modèle de *forme statistique (statistical shape model)* (*Active Shape Model, ASM*) proposé dans [Cootes et autres 1995] offre plus de flexibilité et de robustesse. Les avantages d'employer la prétendue analyse par l'approche de synthèse viennent du fait que la solution *est contrainte* par un modèle statistique *flexible*. Pour prendre en considération les variations de texture, le modèle ASM a été étendu aux *modèles statistiques d'apparence (statistical appearance models)* comprenant un *modèle flexible d'apparence (Flexible Appearance Model, FAM)* [Lanitis et autres 1995] et un *modèle actif d'apparence (Active Appearance Model, AAM)* [Cootes et autres 2001]. Dans [Cootes et autres 2001], l'AAM proposé a combiné un modèle de variation de forme (c.-à-d., ASM) avec un modèle de variation d'apparence de textures à formes normalisées. Un ensemble d'apprentissage de 400 images de visages, chacune étant manuellement étiquetée avec 68 points de bornes limites et approximativement 10.000 valeurs d'intensité prélevées des régions faciales a été utilisé. Le modèle de forme (forme moyenne, matrice orthogonale de projection  $P_s$  et vecteur de projection  $b_s$ ) est généré en

représentant chaque ensemble de bornes limites comme un vecteur et en appliquant *l'analyse en composantes principales (principal component analysis, PCA)* aux données.

Ensuite, après que chaque image échantillon soit déformée de sorte que ses bornes limites assortissent la forme moyenne, l'information de texture peut être prélevée de ce morceau de visage libre de forme. L'application de la PCA à ces données mène à un modèle de texture libre de forme (texture moyenne,  $P_g$  et  $b_g$ ). Pour explorer la corrélation entre la forme et les variations de texture, une troisième PCA est appliquée aux vecteurs concaténés ( $b_s$  et  $b_g$ ) pour obtenir le modèle combiné dans lequel un vecteur  $c$  des paramètres d'apparence contrôle la forme et la texture du modèle. Pour assortir une image donnée avec le modèle, un vecteur optimal de paramètres (les paramètres de déplacement entre la région de visage et le modèle, les paramètres pour l'ajustement linéaire d'intensité et les paramètres d'apparence  $c$ ) sont recherchés en réduisant au minimum la différence entre l'image synthétique et l'image donnée. Après assortiment, on construit un meilleur modèle d'adaptation qui donne les positions de tous les traits faciaux et qui peut être employé pour reconstruire les images originales. La figure 3.2 illustre la procédure de recherche/optimisation pour adapter le modèle à l'image. Pour accélérer le procédé de recherche, une méthode efficace qui exploite les similitudes parmi les optimisations est proposée. Ceci permet à la méthode directe de trouver et d'appliquer les directions de la convergence rapide qui sont appris en mode off line.



**Figure 3.2** Recherche multi-résolutions à partir de positions déplacées en utilisant un modèle de visage [Cootes et autres 2000].

### **3.2.2. Reconnaissance à partir d'images d'intensité:**

Plusieurs méthodes de reconnaissance de visage ont été proposées pendant les 30 dernières années. La reconnaissance de visage est un problème si provocant pourtant intéressant qu'il a attiré les chercheurs de différents domaines: psychologie, reconnaissance de formes, réseaux de neurones, vision par ordinateur et infographie. Il est dû à ce fait que la littérature sur la reconnaissance de visage est vaste et diverse. Souvent, un simple système implique des techniques motivées par différents principes.

L'utilisation d'un mélange de techniques rend difficile la classification de ces systèmes en se basant purement sur les types de techniques utilisées pour la représentation ou la classification de traits. Pour avoir une catégorisation claire et de niveau élevé, on suit une directive suggérée par l'étude psychologique de la façon dont les humains emploient les caractéristiques (traits) holistiques et locales. Spécifiquement, on a la catégorisation suivante :

(1) **Méthodes holistiques d'assortiment (*holistic matching methods*)**: ces méthodes emploient la région entière du visage comme entrée (à l'état brut) à un système de reconnaissance. Une des représentations les plus répandues de la région du visage est celle des *images propres (eigenpictures)* [Kirby et Sirovich 1990 ; Sirovich et Kirby 1987], qui est basé sur *l'analyse en composantes principales (PCA)*.

(2) **Méthodes d'assortiment (ou structurelles) basées sur les traits (*feature-based (structural) matching methods*)**: Typiquement, dans ces méthodes, des traits locaux tels que les yeux, le nez et la bouche sont d'abord extraits et leurs positions et statistiques locales (géométriques et/ou apparence) sont introduits dans un classificateur structurel.

(3) **Méthodes hybrides (*hybrid methods*)**: Justement, comme le système de perception humaine emploie les traits locaux et la région entière du visage pour reconnaître un visage, un système de reconnaissance par machine devrait les employer tous les deux. On peut arguer du fait que ces méthodes pourraient potentiellement offrir le meilleur des deux types de méthodes.

Dans chacune de ces catégories, d'avantage de classification est possible (Tableau 3.3). En utilisant l'analyse en composantes principales (PCA), plusieurs techniques de reconnaissance de visage ont été développées: la technique des faces propres (*eigenfaces*) [Turk et Pentland 1991], qui emploie un *classificateur du voisin le plus proche (nearest neighbor classifier)*, les *méthodes basées sur les lignes de traits (feature-line-based methods)* qui remplacent la distance point à point par la distance entre un point et la ligne du trait liant deux points échantillons stockés [Li et Lu 1999]; la technique des faces de Fisher (*Fisherfaces*) [Belhumeur et autres 1997 ; Liu et Wechsler 2001 ; Swets et Weng 1996b ; Zhao et autres 1998] qui emploie l'analyse linéaire discriminante (*linear/Fisher discriminant analysis, (FLD/LDA)*) [Fisher 1938], les *méthodes bayésiennes*, qui emploient une distance probabiliste métrique [Moghaddam et Pentland 1997] et les *méthodes de machine de vecteur de soutien (Support Vector Machines, SVM)* qui utilisent une machine de vecteur de soutien (SVM) comme classificateur [Phillips 1998]. En utilisant des statistiques d'ordre plus supérieur, *l'analyse en composantes indépendantes (Independent Component Analysis, ICA)* est discutée pour avoir une puissance plus représentative que la PCA et par conséquent peut fournir une performance de reconnaissance meilleure que celle de la PCA [Bartlett et autres 1998]. Pouvant offrir une généralisation potentiellement plus grande par apprentissage, les méthodes se basant sur les réseaux de neurones ont été également appliquées à la reconnaissance du visage. Un exemple est la *méthode du réseau de neurones basée sur la décision probabiliste (Probabilistic Decision-Based Neural Network, PDBNN)* [Lin et autres 1997] et l'autre

---

est la *méthode de la poursuite d'évolution (Evolution Pursuit, EP)* [Liu et Wechsler 2000a].

La plupart des méthodes précédentes appartiennent à la catégorie des méthodes structurelles d'assortiment, en utilisant la largeur de la tête, les distances entre les yeux et les distances des yeux à la bouche, etc.... [Kelly 1970] ou les distances et les angles entre les coins des yeux, les extrema de la bouche, les narines et le menton [Kanade 1973]. Plus récemment, une approche basée sur un mélange de distances (mixture-distance based approach) employant des distances manuellement extraites a été présentée dans [Cox et autres 1996]. Sans trouver les endroits exacts des traits faciaux, les méthodes basées *sur les modèles cachés de Markov* (Hidden Markov Models, HMM) utilisent des bandes de pixels qui couvrent le front, l'oeil, le nez, la bouche et le menton [Nefian et Hayes 1998 ; Samaria 1994 ; Samaria et Young 1994]. [Nefian et Hayes 1998] ont obtenu une performance meilleure que celle de [Samaria 1994] en employant les coefficients de projection de KL (Karhunen-Loeve ou PCA) au lieu des bandes de Pixels. Un des systèmes les plus réussis dans cette catégorie est le système d'assortiment de graphes (graph matching system) [Okada et autres 1998 ; Wiskott et autres 1997], qui est basé sur l'architecture des liens dynamiques (Dinamic Link Architecture DLA) [Buhmann et autres 1990 ; Lades et autres 1993]. En Employant une méthode d'apprentissage non supervisé basée sur une carte à organisation automatique (Self Organizing Map, SOM), un système basé sur un réseau de neurones de convolution (Convolutional Neural Network, CNN) a été développé [Lawrence et autres 1997].

**Tableau 3.3** Catégorisation des techniques de reconnaissance de visage à partir d'images statiques [Zhao et autres 2003].

<b>Approche</b>	<b>Travail représentatif</b>
<b>Méthodes holistiques</b>	
Analyse en composantes principales (PCA)	
Faces propres	Application directe de la PCA [Craw et Cameron 1996 ; Kirby et Sirovich 1990 ; Turk et Pentland 1991]
Faces propres probabilistes	Problème à 2 classes avec mesure de probabilités [Moghaddam et Pentland 1997]
Fisherfaces/subspace LDA	FLD dans l'espace propre [Belhumeur et autres 1997 ; Swets et Weng 1996b ; Zhao et autres 1998]
SVM	Problème à deux classes basée sur SVM [Phillips 1998]
Poursuite d'évolution	Apprentissage amélioré par AG [Liu et Wechsler 2000a]
Lignes des traits	Distance point à ligne [Li et Lu 1999]
ICA	Analyse de traits basée sur ICA [Bartlett et autres 1998]
Autres représentations	
LDA/FLD	LDA/FLD sur image brute [Etemad et Chellappa 1997]
PDBNN	PDBNN [Lin et autres 1997]
<b>Méthodes basées sur les traits</b>	
Méthodes géométriques pures	Anciennes méthodes [Kanade 1973 ; Kelly 1970] Méthodes récentes [Cox et autres 1996 ; Manjunath et autres 1992]
Architecture à liens dynamiques	Méthodes d'assortiment de graphes [Okada et autres 1998 ; Wiskott et autres 1997]
Modèles cachés de Markov	Méthodes HMM [Nefian et Hayes 1998 ; Samaria 1994 ; Samaria et Young 1994]
Réseau de neurones de convolution	Méthodes CNN basées sur l'apprentissage des SOM [Lawrence et autres 1997]
<b>Méthodes hybrides</b>	
Faces propres modulaires	Faces propres et modules propres [Pentland et autres 1994]
LFA hybride	Méthode des traits locaux [Penev et Atick 1996]
à formes normalisées	Modèles flexibles d'apparence [Lanitis et autres 1995]
Basées sur les composantes	Région de la face et les composantes [Huang et autres 2003]

Dans la catégorie des méthodes hybrides, on passera en revue brièvement *la méthode des faces propres modulaires (modular eigenface method)* [Pentland et autres 1994], une représentation hybride basée sur la PCA et *l'analyse des traits locaux (local feature analysis, LFA)* [Penev et Atick 1996], *une méthode basée sur les modèles flexibles d'apparence (flexible appearance model-based method)* [Lanitis et autres 1995] et un développement récent [Huang et autres 2003]. Dans [Pentland et autres 1994], l'utilisation des traits hybrides en combinant des faces propres avec des modules propres est explorée : *yeux propres (eigeneyes)*, *bouche propre (eigenmouth)* et *nez propre (eigennose)*. Bien que les expériences montrent de légères améliorations comparées aux faces propres holistiques ou aux modules propres holistiques basées sur l'assortiment structurel, on croit que ces types de méthodes sont importantes et méritent d'avantage de recherche. Peut-être, beaucoup de problèmes appropriés ont besoin d'être résolu avant que des résultats fructueux puissent être prévus, par exemple, comment arbitrer de façon optimale l'utilisation des traits holistiques et locaux.

Plusieurs types de systèmes ont été appliqués avec succès à la tâche de la reconnaissance de visage, mais ils ont tous quelques avantages et inconvénients. Des arrangements appropriés devraient être choisis en se basant sur les conditions spécifiques d'une tâche donnée. La plupart des systèmes passés en revue ici se concentrent seulement sur la tâche (secondaire) de la reconnaissance, mais d'autres incluent également la détection automatique de visage et l'extraction de traits, faisant d'eux des systèmes entièrement automatiques [Lin et autres 1997 ; Moghaddam et Pentland 1997 ; Wiskott et autres 1997].

### **3.2.2.1. Approches holistiques:**

#### **3.2.2.1.1. L'analyse en composantes principales:**

À partir de la reconstruction réussie, en petite dimension des visages en utilisant les projections de KL/PCA [Kirby et Sirovich 1990 ; Sirovich et Kirby 1987], les images propres ont été l'une des forces d'entraînement principales derrière la représentation de visage, la détection, et la reconnaissance. Il est bien connu qu'il existe des redondances statistiques significatives dans les images naturelles [Ruderman 1994]. Pour une classe limitée d'objets tels que les images de visage qui sont normalisées par rapport à la taille, la translation et la rotation, la redondance est encore plus grande [Penev et Atick 1996 ; Zhao 1999]. Une des meilleures représentations compactes et globales est la KL/PCA, qui décortique les sorties.

Plus spécifiquement, les vecteurs échantillons  $x$  peuvent être écrits comme une combinaison linéaire de la base orthogonale  $\Phi_i$  :

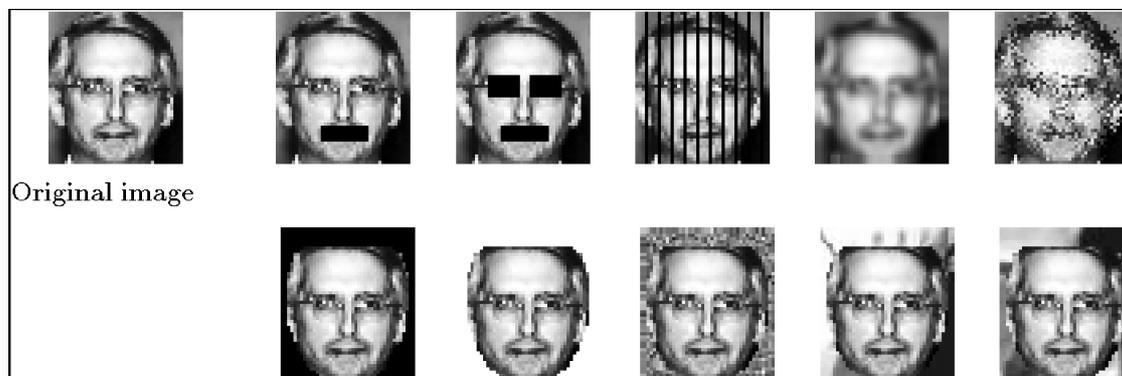
$$x = \sum_{i=1}^n a_i \Phi_i \approx \sum_{i=1}^m a_i \Phi_i \quad (3.1)$$

(typiquement,  $m \ll n$ )

En résolvant le problème des valeurs propres:

$$C \Phi = \Phi \Lambda \quad (3.2)$$

Où  $C$  est la matrice de covariance des entrées  $x$ .



**Figure 3.3** Des images modifiées électroniquement et qui ont été correctement identifiées ([Zhao et autres 2003]).

Un avantage dans l'utilisation de telles représentations est leur sensibilité réduite au bruit. Une partie de ce bruit peut être due à de petites occlusions, aussi longtemps que la structure topologique ne change pas. Par exemple, la bonne exécution sous l'effet du flou, de l'occlusion partielle et des changements d'arrière plan a été démontrée dans beaucoup de systèmes basés sur les images propres comme illustré sur la figure 3.3. Ceci ne devrait pas nous surprendre puisque les images reconstruites par la PCA sont meilleures que les images originales ayant subi des distorsions en termes de leur apparence globale (voir figure 3.4).



**Figure 3.4** Images reconstruites en utilisant 300 coefficients de projection de la PCA pour les images modifiées électroniquement de la figure 3.3. ([Zhao et autres 2003]).

Pour une meilleure approximation des images de visage en dehors de l'ensemble d'apprentissage, l'emploi d'un ensemble d'apprentissage augmenté qui ajoute des images qui sont des reflets miroirs des images originales a permis l'obtention d'une erreur d'approximation plus petite [Kirby et Sirovich 1990]. En utilisant un tel ensemble d'apprentissage augmenté, les images propres sont symétriques ou antisymétriques, avec

les principales images propres (correspondantes aux valeurs propres les plus grandes) étant typiquement symétriques.

La première démonstration vraiment réussie de la reconnaissance de visage par des machines a été faite dans [Turk et Pentland 1991] en employant des images propres (également connues sous le nom de faces ou visages propres) pour la détection et la reconnaissance de visage. Etant données les faces propres, chaque visage dans la base de données peut être représenté comme un vecteur de poids; les poids sont obtenus en projetant l'image dans l'espace des faces propres par une simple opération de produit scalaire. Quand une nouvelle image de test dont l'identification est exigée est donnée, la nouvelle image est également représentée par son vecteur de poids. L'identification de l'image de test est faite en localisant l'image dans la base de données dont les poids sont les plus proches des poids de l'image de test. En employant l'observation que la projection d'une image de visage et d'une image de non visage sont habituellement différentes, une méthode de détecter la présence d'un visage dans une image donnée est obtenue. La méthode a été démontrée en utilisant une base de données de 2500 images de visage de 16 personnes (individus), dans toutes les combinaisons de trois orientations de visage, de trois tailles de visage et de trois conditions d'illumination.

En utilisant une mesure probabiliste de similitude au lieu de la distance euclidienne simple utilisée avec les faces propres [Turk et Pentland 1991], l'approche standard des faces propres a été étendue [Moghaddam et Pentland 1997] à une approche bayésienne. Pratiquement, l'inconvénient principal d'une méthode bayésienne est la nécessité d'estimer des distributions de probabilité dans un espace dimensionnel élevé à partir d'un nombre très limité d'échantillons d'apprentissage par classe. Pour éviter ce problème, un problème beaucoup plus simple de deux classes a été créé à partir du problème multi classes en employant une mesure de similitude basée sur une analyse bayésienne des différences d'images. Deux classes mutuellement exclusives ont été définies:  $\Omega_I$  représente les variations intra personnelles entre les image d'un même individu et  $\Omega_E$  qui représente les variations inter personnelles dues aux différences d'identités. En assumant que chacune des deux classes est distribué suivant la loi Gaussienne, les distributions de probabilités  $P(\Delta/\Omega_I)$  et  $P(\Delta/\Omega_E)$  sont estimées pour une différence d'intensité donnée  $\Delta = I_1 - I_2$ . Etant données ces deux distributions de probabilités, deux images de visage sont considérées comme appartenant au même individu si  $P(\Delta/\Omega_I) > P(\Delta/\Omega_E)$ .

Une grande amélioration de performance de cette méthode d'assortiment probabiliste par rapport à la méthode standard d'assortiment des visages propres avec le voisin le plus proche (nearest neighbor) a été rapportée en utilisant de grandes bases de données de visage incluant la base de données FERET [Phillips et autres 2000]. En [Moghaddam et Pentland 1997], une technique efficace pour l'estimation de la densité de probabilité a été propose en décomposant l'espace d'entrée en deux sous-espaces mutuellement exclusifs: le sous-espace principal  $F$  et son sous-espace orthogonal  $F^\perp$  (une idée semblable a été explorée dans [Sung et Poggio 1997]). Seules les covariances

dans le sous-espace principal sont estimées pour l'usage dans la distance de Mahalanobis [Fukunaga 1989].

Des résultats expérimentaux ont été rapportés en utilisant différentes dimensionnalités  $M_I$  et  $M_E$  de sous-espaces pour  $\Omega_I$  et  $\Omega_E$ . Par exemple,  $M_I = 10$  et  $M_E = 30$  ont été employées pour les tests internes, alors que  $M_I = M_E = 125$  étaient employées pour le test de FERET. Sur la figure 3.5, les prétendues faces propres duales dont l'apprentissage a été effectué séparément sur des échantillons de  $\Omega_I$  et de  $\Omega_E$  sont dessinées avec les faces propres standard. Tandis que les faces propres inters personnelles ressemblent plus aux faces propres standard que les faces intra personnelles, les faces propres intra personnelles représentent les variations subtiles qui sont dues dans la plupart du temps à l'expression faciale et à l'éclairage, suggérant qu'elles soient plus critiques pour l'identification [Moghaddam et Pentland 1997].

Les systèmes de reconnaissance de visage utilisant la LDA/FLD ont également été très réussis [Belhumeur et autres 1997 ; Etemad et Chellappa 1997 ; Swets et Weng 1996b ; Zhao et autres 1998 ; Zhao et autres 1999]. L'apprentissage de la LDA est effectué par l'intermédiaire de l'analyse de la matrice de dispersion [Fukunaga 1989]. Pour un problème à  $M$  classes, les matrices de dispersion intra classes et inter classes  $S_\omega$  et  $S_b$  sont calculées comme suit:

$$S_\omega = \sum_{i=1}^M \Pr(\omega_i) C_i \quad (3.3)$$

$$S_b = \sum_{i=1}^M \Pr(\omega_i) (m_i - m_0)(m_i - m_0)^T \quad (3.4)$$

Où  $\Pr(\omega_i)$  est la probabilité a priori de la classe et est habituellement remplacée par  $1/M$  dans la pratique avec la supposition de probabilités a priori égales. Ici,  $S_\omega$  est la matrice de dispersion intra classes qui mesure la dispersion moyenne  $C_i$  des vecteurs échantillons  $x$  des différentes classes  $\omega_i$  autour de leurs moyennes respectives  $m_i$  :

$$C_i = E[(x(\omega) - m_i)(x(\omega) - m_i)^T \mid \omega = \omega_i] \quad (3.5)$$

De même,  $S_b$  est la matrice de dispersion inter classes, représentant la dispersion moyenne des vecteurs des moyennes conditionnelles  $m_i$  autour du vecteur de la moyenne totale  $m_0$ . Une mesure généralement utilisée pour quantifier la puissance de discrimination est le rapport du déterminant de la matrice de dispersion inter classes et du déterminant de la matrice de dispersion intra classes des échantillons projetés:

$$\mathfrak{S}(T) = |T^T S_b T| / |T^T S_\omega T|. \quad (3.6)$$

La matrice de projection optimale  $W$  qui maximise  $\mathfrak{J}(T)$  est obtenue en résolvant le problème des valeurs propres généralisées:

$$S_b W = S_w W \Lambda_W \quad (3.7)$$

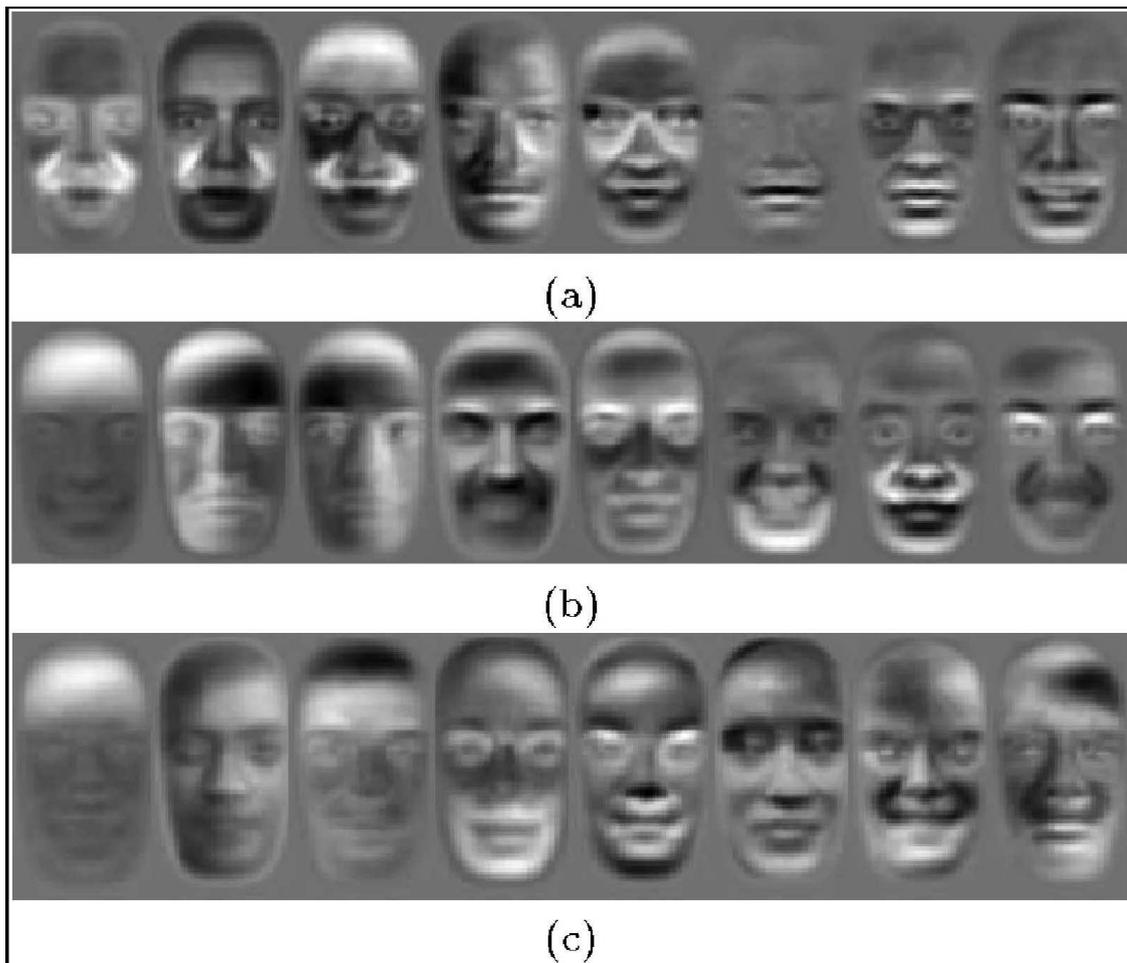
Il est utile de faire des comparaisons parmi les prétendus algorithmes (linéaires) de projection. Ici, on illustre la comparaison entre les faces propres et les faces de Fisher. Des comparaisons semblables peuvent être faites pour d'autres méthodes, par exemple, les méthodes de projection de la ICA. Dans tous ces algorithmes de projection, la classification est effectuée en:

(1) projetant l'entrée  $x$  dans un sous-espace par l'intermédiaire d'une matrice de projection (ou des bases)  $P_{roj}$  ( $P_{roj}$  est égale à:  $\Phi$  pour les faces propres,  $W$  pour les faces de Fisher avec projection pure de la LDA,  $W \Phi$  pour les faces de Fisher avec des projections séquentielles de la PCA et de la LDA), ces trois bases sont montrées (pour une comparaison visuelle) en figure 3.6):

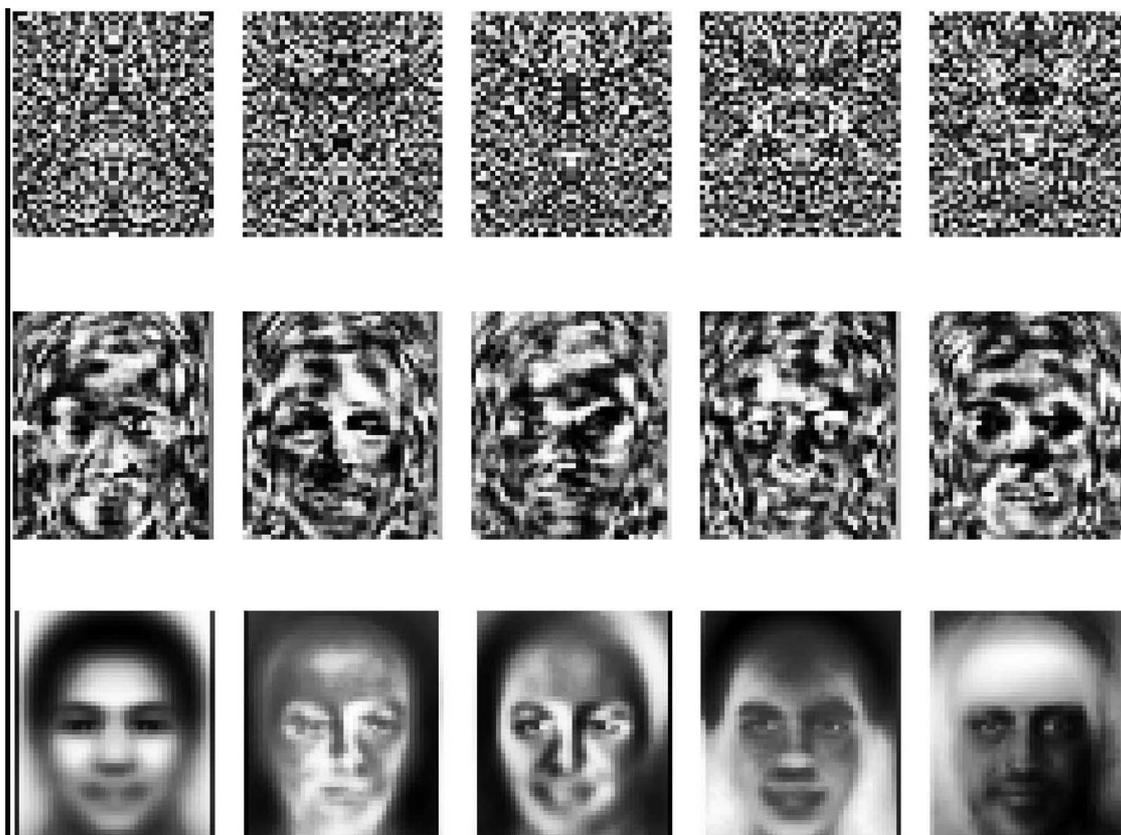
$$z = P_{roj} x \quad (3.8)$$

(2) comparant le vecteur  $z$  des coefficients de projection de l'entrée à tous les vecteurs préenregistrés de projection des classes connues (étiquetées) pour déterminer l'étiquette de classe d'entrée. La comparaison de vecteurs varie d'une implémentation à l'autre et peut nettement influencer la performance du système [Moon et Phillips 2001]. Par exemple, les algorithmes de la PCA peuvent employer la distance angulaire ou la distance Euclidienne (pondérée ou non pondérée) entre deux vecteurs de projection. Pour les algorithmes de la LDA, la distance peut être aussi pondérée ou non pondérée.

Dans [Swets et Weng 1996b], l'analyse discriminante des faces propres est appliquée dans un système de récupération d'images pour déterminer non seulement la classe (visage humain ou non visage), mais également les individus dans la classe du visage. En utilisant l'apprentissage de structure d'arbre (tree-structure learning), les projections de l'espace propre (eigenspace) et de la LDA sont appliquées récursivement à des ensembles d'échantillons de plus en plus petits. Un tel partitionnement récursif est effectué pour chaque noeud jusqu'à ce que les échantillons assignés au noeud appartiennent à une même classe. Des expériences sur cette approche ont été rapportées dans [Swets et Weng 1996]. Un ensemble de 800 images a été employé pour l'apprentissage, l'ensemble d'apprentissage provient de 42 classes, desquelles les visages humains appartiennent à une seule classe. Dans l'unique classe de visage, 356 individus étaient inclus et distingués. Les résultats des tests sur des images n'appartenant pas à l'ensemble d'apprentissage étaient de 91% de réussite pour 78 images de visage et de 87% de réussite pour 38 images de non visage.



**Figure 3.5** comparaison entre les faces propres duales et les faces propres standard: (a) intra personnelles, (b) inter personnelles et (c) standard ([Moghaddam et Pentland 1997]).



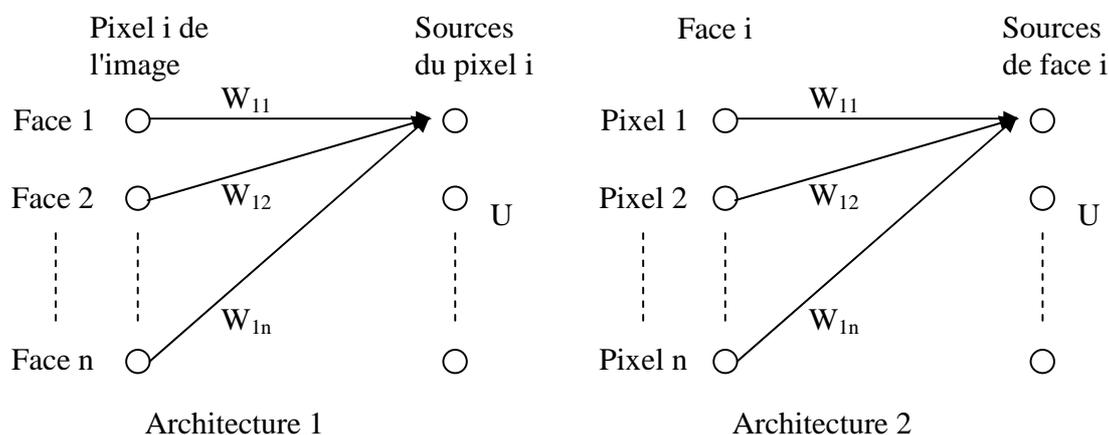
**Figure 3.6** Différentes bases de projection construites à partir d'un ensemble de 444 individus, où l'ensemble est augmenté en ajoutant du bruit et du reflet. La première rangée montre les cinq premières images de la base  $W$  de la LDA pure, la seconde rangée montre les cinq premières images de la base  $W \Phi$  de la "subspace LDA", la face moyenne et les quatre premières faces propres  $\Phi$  sont montrée dans la troisième rangée ([Zhao et autres 1998]).

Une analyse comparative de performance a été effectuée dans [Belhumeur et autres 1997]. Quatre méthodes ont été comparées dans cet article: (1) une méthode basée sur la corrélation, (2) une variante de la méthode des sous-espaces linéaires suggérée en [Shashua 1994], (3) une méthode des faces propres [Turk et Pentland 1991] et (4) une méthode des faces de Fisher qui emploie la projection des sous-espaces avant la projection de la LDA pour éviter la singularité possible dans  $S_w$  comme dans [Swets et Weng 1996b]. Les expériences ont été effectuées sur une base de données de 500 images créées par [Hallinan 1994] et une base de données de 176 images créées chez Yale. Les résultats de ces expériences ont prouvé que la méthode des faces de Fisher donne une performance sensiblement meilleure que les trois autres méthodes. Cependant, aucune réclamation n'a été faite au sujet des performances relatives de ces algorithmes sur de plus grandes bases de données.

Pour améliorer les performances des systèmes basés sur la LDA, Un système régularisé de sous-espaces de la LDA qui unifie PCA et LDA a été propose en [Zhao 1999] et en [Zhao et autres 1998]. De bonnes capacités de généralisation de ce système ont été démontrées par des expériences qui effectuent des tests sur de nouvelles classes (personnes ou individus) sans refaire l'apprentissage des bases  $\Phi$  de la PCA et parfois des bases  $W$  de la LDA. Tandis que la raison de ne pas refaire l'apprentissage de la PCA est évidente, il est intéressant d'examiner les possibilités adaptatives du système en fixant les bases de la LDA quand des images de nouvelles classes sont ajoutées (Ceci se comprend parce que la classification finale est effectuée dans l'espace de projection  $z$  par comparaison avec des vecteurs de projection préenregistrés). Le sous-espace fixé de la PCA de dimensionnalité 300 a été appris à partir d'un grand nombre d'échantillons. Un ensemble augmenté de 4056 images (de vue frontale dans la plupart du temps) construites des 1078 images originales FERET de 444 individus en ajoutant du bruit et du reflet a été employé dans [Zhao et autres 1998]. Au moins une des trois caractéristiques suivantes sépare ce système des autres systèmes basés sur la LDA: (1) le choix unique de la dimension universelle du sous-espace de visage, (2) l'utilisation d'une mesure pondérée de distance et (3) un procédé régularisé qui modifie la matrice de dispersion intra classes  $S_{\omega}$ . Les auteurs ont choisi la dimensionnalité du sous-espace universel de visage en se basant sur les caractéristiques des vecteurs propres (comme un visage ou non) au lieu des valeurs propres [Zhao et autres 1998], comme c'est généralement fait. Plus tard il a été conclu dans [Penev et Sirovich 2000] que la dimensionnalité globale du sous-espace de visage est de l'ordre de 400 pour de grandes bases de données de 5000 images. Une distance métrique pondérée dans l'espace de projection  $z$  a été employée pour améliorer la performance [Zhao 1999] (Les métriques pondérées ont été utilisées aussi dans l'approche de la LDA pure [Etemad et Chellappa 1997] et dans l'approche des FLD améliorés (enhanced FLD, EFM) dans [Liu et Wechsler 2000b]). Finalement, l'apprentissage de la LDA a été régularisée en modifiant la matrice  $S_{\omega}$  en  $S_{\omega} + \delta I$ , où  $\delta$  est un nombre positif relativement petit. Faire ceci résout un problème numérique quand  $S_{\omega}$  est presque singulière. Dans le cas extrême où seulement un échantillon par classe est disponible, cette régularisation transforme le problème de la LDA en problème standard de PCA avec  $S_b$  étant la matrice de covariance  $C$ . En appliquant cette approche, sans refaire l'apprentissage de la base de la LDA à un ensemble de test de 46 individus desquels 24 ont été appris et 22 n'ont pas été appris (un total de 115 images comprenant 19 images non apprises de vues non frontales), les auteurs ont rapporté les performances suivantes basées sur une base de données d'apprentissage de 738 images de vues frontales seulement: 85.2% pour toutes les images et 95.1% pour les vues frontales.

Une représentation adaptative basée de la poursuite d'évolution (EP) et son application à la reconnaissance de visage ont été présentées dans [Liu et Wechsler 2000a]. Dans l'analogie aux méthodes de poursuite de projection, la EP cherche à apprendre une base optimale pour les deux fins de la compression de données et de la classification des formes. Afin d'augmenter les capacités de généralisation de la EP, un équilibre est cherché entre la minimisation du risque empirique rencontré pendant l'apprentissage et le rétrécissement de l'intervalle de confiance pour réduire le risque garanti pendant les futurs tests sur de nouvelles données [Vapnik 1995]. La EP met en

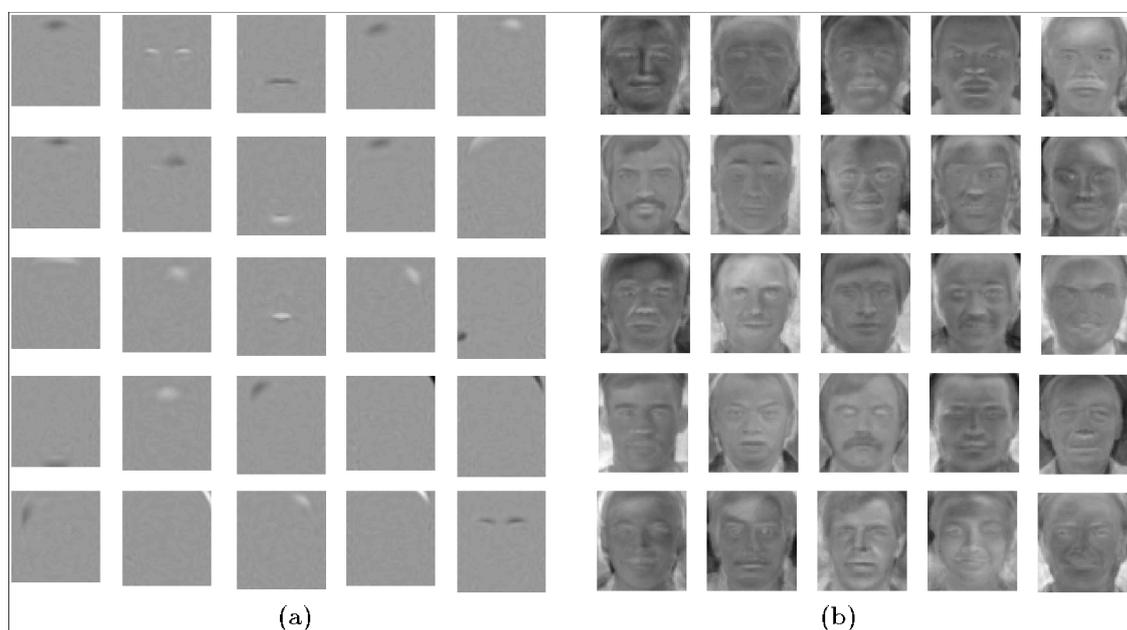
application des stratégies caractéristiques des algorithmes génétiques (GAs) pour rechercher l'espace des solutions possibles pour déterminer la base optimale. La EP commence par projeter les données originales dans un espace de dimension basse blanchi (whitened) de la PCA. Des rotations aléatoires dirigées des vecteurs de base dans cet espace sont alors recherchées par les algorithmes génétiques, où l'évolution est contrôlée par une fonction de performance définie en termes de précision de la performance (risque empirique) et de séparation de classes (intervalle de confiance). La praticabilité de cette méthode a été démontrée pour la reconnaissance de visage, où le grand nombre de bases possibles exige un algorithme gourmand de recherche. La tâche particulière de la reconnaissance de visage implique 1107 images de visages de vue frontale FERET de 369 sujets, il y avait trois images de vue frontale pour chaque sujet, deux pour l'apprentissage et le reste pour le test. Les auteurs ont rapportés une amélioration de la performance de la reconnaissance de visage par rapport aux faces propres [Turk et Pentland 1991] et de meilleures possibilités de généralisation que les faces de Fisher [Belhumeur et autres 1997].



**Figure 3.7** Deux architectures pour appliquer la ICA sur les images. A gauche: architecture pour trouver les images de bases statistiquement indépendantes. L'application d'une séparation de sources aux images de visage produit des images indépendantes sans les lignes de  $U$ . A droite: architecture pour trouver un code factoriel. L'application d'une séparation de sources aux pixels produit un code factoriel dans les colonnes de la matrice de sortie  $U$  ([Bartlett et autres 1998]).

Basé sur l'argument que pour des tâches telles que la reconnaissance de visage, une grande partie de l'information importante est contenu dans des statistiques d'ordre élevé, il a été proposé dans [Bartlett et autres 1998] d'employer la ICA pour extraire les traits (caractéristiques) pour la reconnaissance de visage. L'analyse en composantes indépendantes est une généralisation de l'analyse en composantes principales, qui décortique les moments d'ordre élevé de l'entrée en plus des moments du second ordre. Deux architectures ont été proposées pour la reconnaissance de visage (voir figure 3.7): la première est utilisée pour trouver un ensemble d'images sources statistiquement indépendantes qui peuvent être considérées comme des traits d'image (caractéristiques) indépendants pour un ensemble donné d'images d'apprentissage [Bell et Sejnowski 1995]

et la seconde est utilisée pour trouver les filtres d'image qui produisent des sorties statistiquement indépendantes (une méthode de code factoriel) [Bell et Sejnowski 1997]. Dans les deux architectures, la PCA est employée d'abord pour réduire la dimensionnalité de la taille originale des images ( $60 \times 50$ ). La ICA est exécuté sur les 200 premiers vecteurs propres dans la première architecture et est effectué sur les 200 premiers coefficients de projection de la PCA dans la deuxième architecture. Les auteurs ont rapporté une amélioration des performances des deux architectures par rapport à la méthode des faces propres dans le scénario suivant: un sous-ensemble de FERET contenant 425 individus a été utilisé, toutes les vues frontales (une par classe) ont été utilisées pour l'apprentissage et les vues frontales restantes (jusqu'à trois) pour le test. Des images de base des deux architectures sont montrées sur la figure 3.8 avec les faces propres correspondantes.



**Figure 3.8** Comparaison des images de base en utilisant deux architectures pour appliquer la ICA: (a) 25 composantes indépendantes de l'architecture 1, (b) 25 composantes indépendantes de l'architecture 2 ([Bartlett et autres 1998]).

#### **3.2.2.1.2. Les autres représentations:**

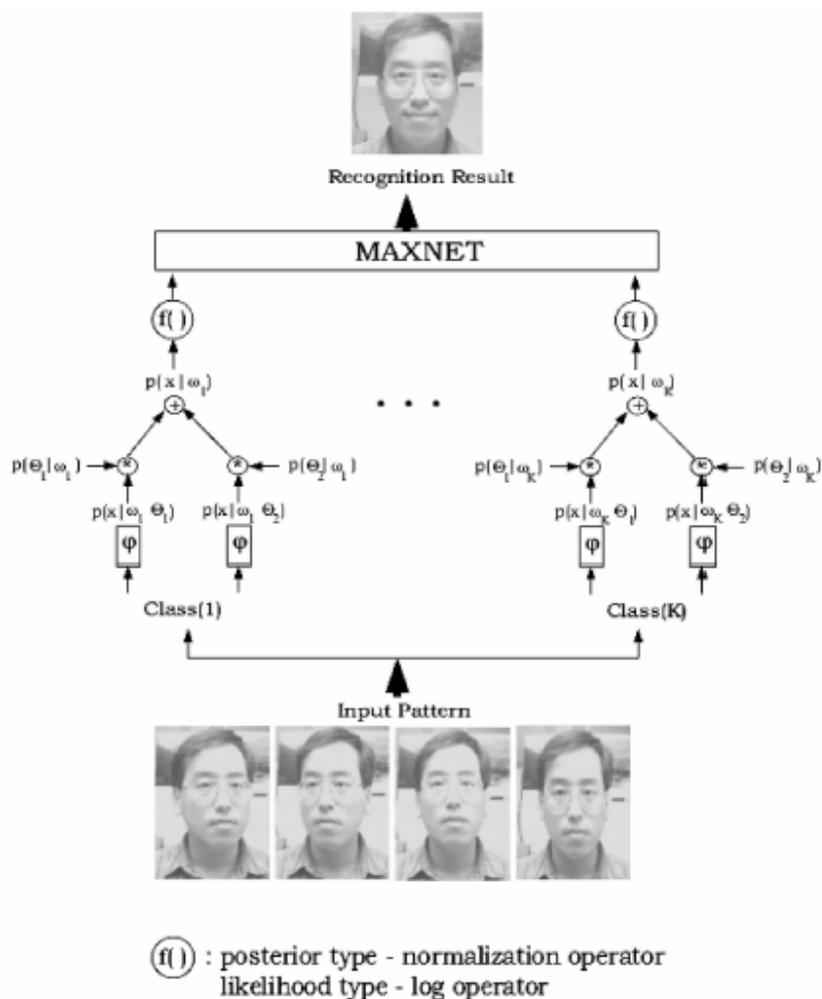
En plus de la représentation populaire de la PCA et de ses dérivés telles que la ICA et la EP, d'autres traits (caractéristiques) ont été également utilisés comme des intensités et des bords à l'état brut.

Un système entièrement automatique de détection/reconnaissance de visage basé sur un réseau de neurones est rapporté en [Lin et autres 1997]. Le système proposé est basé sur un réseau de neurones basé sur une décision probabiliste (PDBNN, un DBNN étendu [Kung et Taur 1995]) qui se compose de trois modules: un détecteur de visage, un localisateur d'oeil et un système de reconnaissance de visage. À la différence de la plupart

des méthodes, les régions faciales contiennent les sourcils, les yeux, et le nez, mais pas la bouche (une telle représentation a été utilisée aussi dans [Kirby et Sirovich 1990]). La raison d'utiliser seulement le visage supérieur est d'établir un système robuste qui exclut l'influence des variations faciales dues aux expressions qui causent le mouvement autour de la bouche. Pour améliorer la robustesse, les images segmentées de la région faciale sont d'abord traitées pour produire deux traits à une résolution réduite de 14x10: traits d'intensité normalisés et traits de bord, tous les deux dans la gamme [0,1]. Ces traits sont introduits dans deux PDBNNs et le résultat final de reconnaissance est la fusion des sorties de ces deux PDBNNs. Une caractéristique unique des PDBNNs et des DBNNs est leur structure modulaire. C'est-à-dire, pour que chaque classe/personne soit reconnue, le PDBNN/DBNN consacre un de ses sous réseaux à la représentation de cette personne particulière, comme illustré sur la figure 3.9. Une telle structure de "une classe par réseau (One-Class-in-One-Network, OCON)" a certains avantages par rapport à la structure de "toutes les classes dans un réseau (All-Classes-in-One-Network, ACON)" qui est adoptée par le perceptron multicouche conventionnel (MultiLayer Perceptron, MLP). Dans la structure ACON, toutes les classes sont réunies dans un super réseau, ainsi un grand nombre d'unités cachées sont nécessaires et la convergence est lente. D'autre part, la structure OCON se compose de sous réseaux qui se composent d'un nombre restreint d'unités cachées, par conséquent, elle converge non seulement plus rapidement mais a également de meilleures possibilités de généralisation. Comparé à la plupart des systèmes de reconnaissance multi classes qui emploient une fonction de discrimination entre deux classes quelconques, le PDBNN a un taux de fausse acceptante/rejection inférieur parce qu'il utilise la description de la densité complète pour chaque classe. En outre, cette architecture est bénéfique pour une implémentation matérielle telle que les architectures distribuées. Cependant, la manière d'estimer avec exactitude, les fonctions de densité complète pour les classes quand il y a seulement des nombres limités d'échantillons n'est pas claire. De plus, le système pourrait avoir des problèmes quand le nombre de classes croît exponentiellement.

### **3.2.2.2. Les approches d'assortiment structurelles basées sur les traits (Feature-Based Structural Matching Methods):**

On a proposé beaucoup de méthodes dans la catégorie de l'assortiment structurel, y compris beaucoup de méthodes anciennes basées sur la géométrie des traits locaux [Kanade 1973 ; Kelly 1970] aussi bien que les méthodes HMM unidimensionnelles [Samaria et Young 1994] et pseudo-bisimensionnelles [Samaria 1994]. Un des plus réussis de ces systèmes est le système d'assortiment de groupes de graphes élastiques (Elastic Bunch Graph Matching, EBGM) [Okada et autres 1998 ; Wiskott et autres 1997], qui est basé sur DLA [Buhmann et autres 1990 ; Lades et autres 1993]. Les *ondelettes* (*wavelets*), particulièrement les *ondelettes de Gabor* (*Gabor wavelets*), jouent un rôle de module pour la représentation faciale dans ces méthodes d'assortiment de graphes. Une représentation locale typique de traits se compose des coefficients des ondelettes pour différentes tailles et rotations basé sur des bases fixes des ondelettes (appelées *les jets* dans [Okada et autres 1998]). Ces coefficients localement estimés des ondelettes sont robustes aux changements de l'illumination, à la translation, à la distorsion, à la rotation et à la taille.



**Figure 3.9** Structure du PDBNN. Chaque sous réseau de classe est conçu pour reconnaître une seule personne. Toutes les pondérations des réseaux sont en format probabiliste ([Lin et autres 1997]).

La fonction de Gabor 2D la plus simple et sa transformée de Fourier sont:

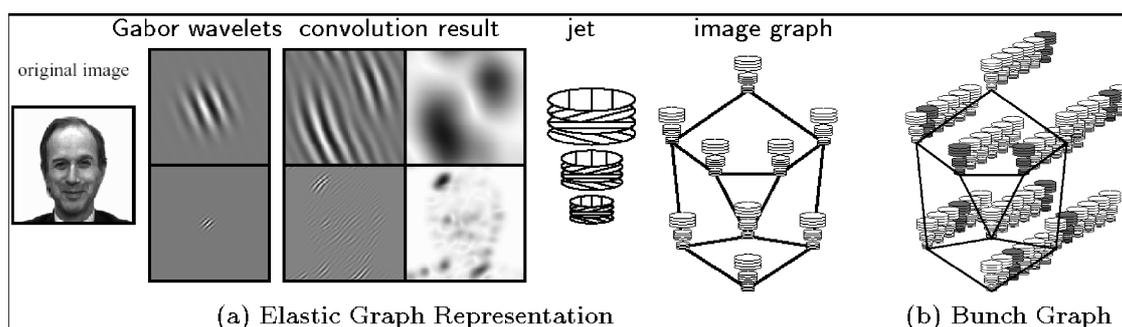
$$g(x, y : u_0, v_0) = e^{-\left[ \frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} \right] + 2\pi i [u_0 x + v_0 y]}, \quad (3.9)$$

$$G(u, v) = e^{-2\pi^2 (\sigma_x^2 (u - u_0)^2 + \sigma_y^2 (v - v_0)^2)}$$

Où  $\sigma_x$  et  $\sigma_y$  représentent les largeurs spatiales des Gaussiennes et  $(u_0, v_0)$  est la fréquence de la sinusoïde complexe.

Les DLAs ont tenté de résoudre certains des problèmes conceptuels des réseaux de neurones artificiels conventionnels, le plus important de ces derniers étant la représentation des rapports syntactiques dans les réseaux de neurones. Les DLAs utilisent *la plasticité synaptique (synaptic plasticity)* et peuvent former des ensembles de neurones groupés dans des graphes structurés tout en maintenant les avantages des systèmes neuronaux. [Buhmann et autres 1990] et [Lades et autres 1993] utilisaient les ondelettes de Gabor (figure 3.10 (a)) comme traits. Comme décrit dans le mécanisme de base des DLAs de [Lades et autres 1993], en plus du paramètre de connexion  $T_{ij}$  entre deux neurones  $(i, j)$  qui est une variable dynamique  $J_{ij}$ . Seules les variables  $J$  jouent les rôles des poids synaptiques pour la transmission de signal. Les paramètres  $T$  agissent simplement pour contraindre les variables  $J$ , par exemple,  $0 \leq J_{ij} \leq T_{ij}$ . Les paramètres  $T$  peuvent être changés lentement par plasticité synaptique à long terme. Les poids  $J_{ij}$  sont sujets à une modification rapide et sont contrôlés par les corrélations de signal entre les neurones  $i$  et  $j$ . Les corrélations négatives de signal mènent à une diminution et les corrélations positives de signal mènent à une augmentation de  $J_{ij}$ . En l'absence de corrélation,  $J_{ij}$  revient lentement à un état de repos, une fraction fixe de  $T_{ij}$ . Chaque image stockée est formée en sélectionnant une grille rectangulaire de points comme noeuds du graphe. La grille est convenablement positionnée au-dessus de l'image et est stockée avec chaque jet localement déterminé des points de la grille (figure 3.10 (a)) et sert pour représenter les classes de formes. La reconnaissance d'une nouvelle image a lieu en transformant l'image en grille de jets et en assortissant tous les graphes des modèles stockés à l'image. La conformation de la DLA est faite en établissant et en modifiant dynamiquement des liens entre les sommets dans le domaine du modèle.

L'architecture DLA a été récemment prolongée à l'assortiment de groupes de graphes élastiques [Wiskott et autres 1997] (figure 3.10). C'est semblable au graphe décrit ci-dessus, mais au lieu d'attacher seulement un jet simple à chaque noeud, les auteurs ont attaché un ensemble de jets (appelés *la représentation des groupes de graphes*, voir figure 3.10 (b)), chacun étant dérivé d'une image de visage différente. Pour gérer le problème de la variation de pose, la pose du visage est d'abord déterminée en utilisant l'information à priori de classe [Kruger et autres 1997] et les transformations des jets sous la variation de pose sont apprises [Maurer et Malsburg 1996a]. Des systèmes basés sur l'approche EBGGM ont été appliqués à la détection et à l'extraction du visage, à l'estimation de la pose, à la classification du genre, à la reconnaissance basée sur le croquis d'image et à la reconnaissance générale d'objets. Le succès du système EBGGM peut être dû à sa ressemblance au système visuel humain [Biederman et Kalocsai 1998].



**Figure 3.10** La représentation des groupes de graphes des visages utilisés dans l'assortiment de graphes élastiques ([Wiskott et autres 1997]).

### **3.2.2.3. Les approches hybrides:**

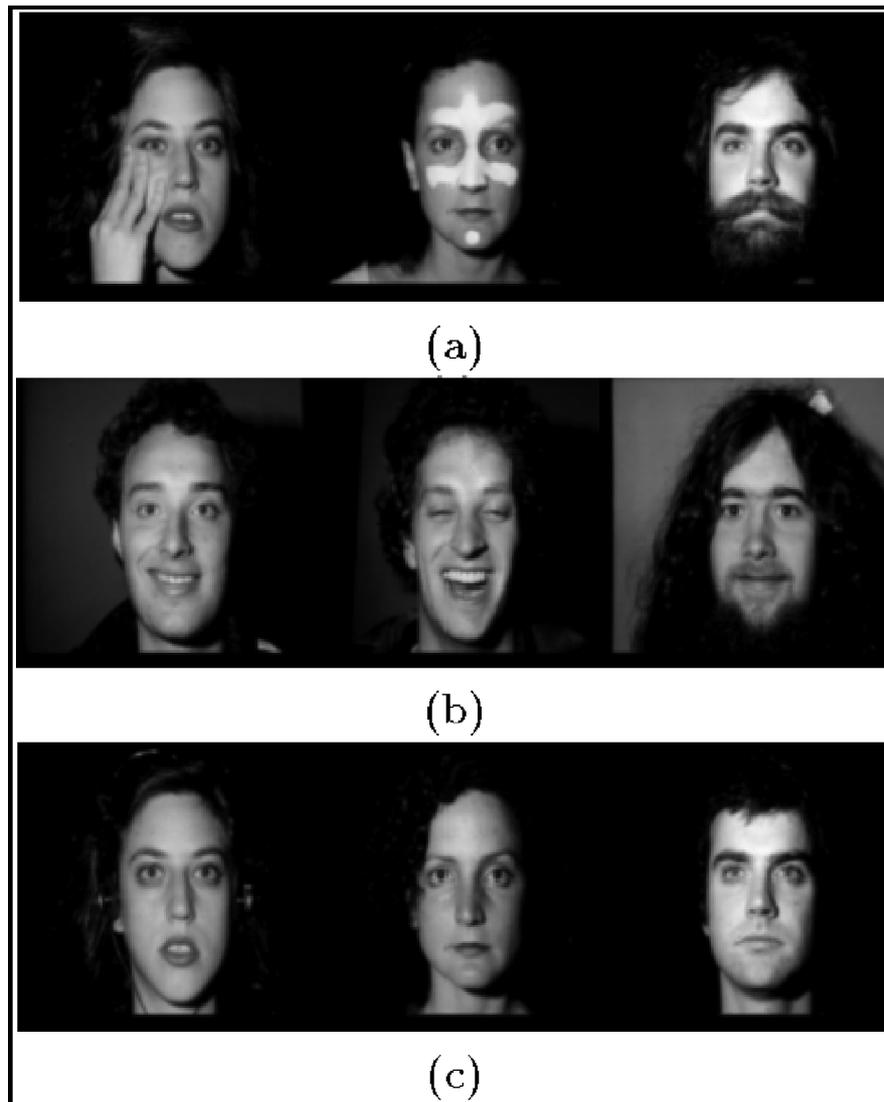
Les approches hybrides emploient des traits holistiques et locaux. Par exemple, l'approche des faces propres modulaires [Pentland et autres 1994] utilise les faces propres globales et les traits propres locaux.

Dans [Pentland et autres 1994], les capacités du système précédent [Turk et Pentland 1991] ont été prolongées dans plusieurs directions. Dans des applications de photographie, une vue frontale et une vue de côté sont généralement disponibles, dans d'autres applications, plus de deux vues peuvent être appropriées. On peut adopter deux approches pour manipuler les images de vues multiples. La première approche rassemble toutes les images et construit un ensemble de faces propres qui représente toutes les images de toutes les vues. L'autre approche utilise des faces propres séparées pour différentes vues, de sorte que la collection d'images prises de chaque vue ait son propre espace propre. La deuxième approche, connue sous le nom *des espaces propres basés sur les vues (view-based eigenspaces)* est la plus performante des deux méthodes.

Le concept des faces propres peut être étendu aux traits propres, tels que les *yeux propres (eigeneyes)*, la *bouche propre (eigenmouth)*, etc.... En utilisant un ensemble limité d'images (45 personnes, deux vues par personne, avec différentes expressions faciales telles que la neutralité contre le sourire), la performance de la reconnaissance en fonction du nombre de vecteurs propres a été mesurée pour les faces propres seules et la représentation combinée. Pour les espaces de petits ordres, les traits propres sont meilleures en termes de performance que les faces propres [Pentland et autres 1994], quand l'ensemble combine a été employé, seulement une amélioration marginale a été obtenue. Ces expériences soutiennent la réclamation que les mécanismes basés sur les traits peuvent être utiles quand des variations brutes sont présentes dans les images d'entrée (figure 3.11).

Il a été argumenté que les systèmes pratiques devraient employer un hybride de PCA et de LFA (annexe B dans [Penev et Atick 1996]). Une telle vue a été longtemps soutenue dans la communauté de la psychologie [Bruce 1988]. Il semble mieux d'estimer les modes propres/faces propres (eigenmodes/eigenfaces) qui ont de grandes valeurs

propres (et ainsi, elles sont plus robustes contre le bruit), alors que pour estimer les modes propres (eigenmodes) de grand ordre, il vaut mieux utiliser la LFA. Pour soutenir ce point, il a été argumenté en [Penev et Atick 1996] que les principales images propres (eigenpictures) sont des filtres globaux, intégrant ou lisseurs qui sont efficaces dans la suppression du bruit, alors que les modes de grand ordre sont des filtres différenciateurs qui sont susceptibles d'amplifier le bruit.



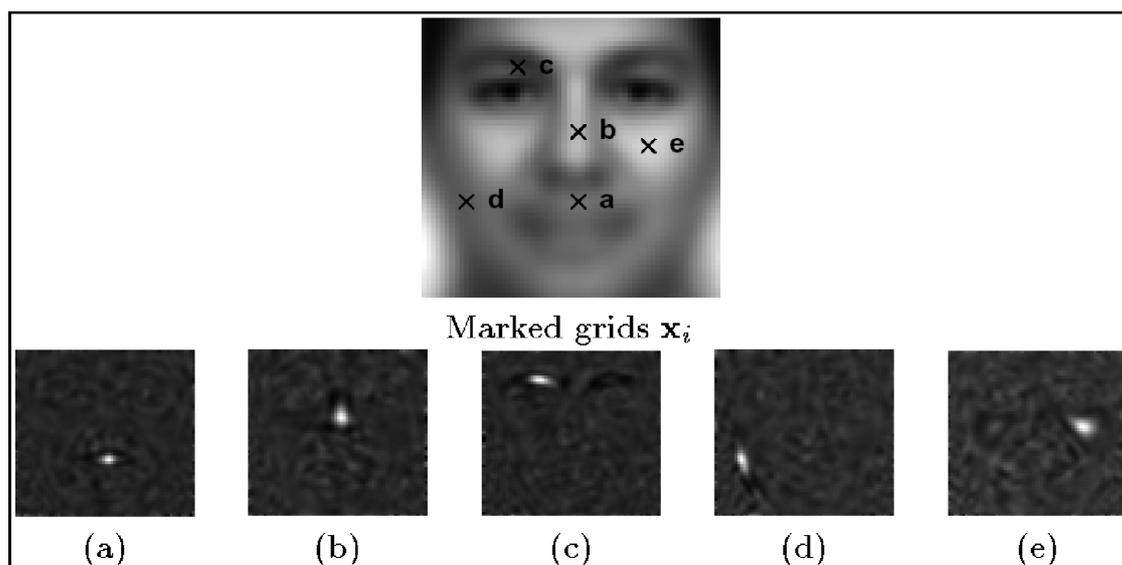
**Figure 3.11** Comparaison de l'assortiment: (a) les vues de test, (b) l'assortiment des faces propres, (c) l'assortiment des traits propres ([Pentland et autres 1994]).

La LFA est une méthode intéressante d'analyse de traits, inspirée biologiquement [Penev et Atick 1996]. Sa motivation biologique vient du fait que malgré une rangée énorme de récepteurs (plus de six millions de cônes) qui existent dans la rétine humaine, seulement une petite fraction est en activité, correspondant aux objets/signaux naturels qui sont statistiquement redondants [Ruderman 1994]. De l'activité de ces récepteurs éparpillés, le cerveau doit découvrir où et quels objets sont dans le champ visuel et récupérer leurs attributs. En conséquence, on cherche à représenter les objets/signaux naturels dans un sous-espace de dimensionnalité réduite en trouvant un paramétrage approprié. Pour une classe limitée d'objets tels que les visages qui sont correctement alignés et de taille normalisée, ceci suggère qu'une dimensionnalité plus réduite puisse être prévue [Penev et Atick 1996]. Un bon exemple est l'utilisation réussie de l'expansion tronquée de la PCA pour approximer les images frontales de visage dans un sous-espace linéaire [Kirby et Sirovich 1990 ; Sirovich et Kirby 1987].

En allant plus loin, la région entière du visage stimule un tableau 2D plein de récepteurs, dont chacun correspond à un endroit dans le visage, mais certains de ces récepteurs peuvent être inactifs. Pour explorer cette redondance, la LFA est utilisée pour extraire les traits locaux topographiques à partir des modes globaux de la PCA. À la différence des noyaux (kernels)  $\Phi_i$  de la PCA, qui ne contiennent aucune information topographique (leurs appuis se prolongent au-dessus de la grille entière des images), les noyaux  $K(x_i, y)$  de la LFA (figure 3.12) à des grilles choisies  $x_i$  ont un appui local. La recherche du meilleur ensemble topographique de grilles éparpillées  $\{x_0\}$  basé sur l'erreur de reconstruction s'appelle *l'éparpillement (sparsification)* et est décrit dans [Penev et Atick 1996]. Deux points intéressants sont démontrés dans cet article: (1) en utilisant le même nombre de noyaux, la qualité perceptuelle de reconstruction de la LFA basée sur l'ensemble optimal de grilles est meilleure que celle de la PCA ; l'erreur quadratique moyenne est de 227 et de 184 pour une entrée particulière; (2) le maintien du deuxième modèle propre de la PCA dans la reconstruction de la LFA réduit l'erreur quadratique moyenne à 152, suggérant une utilisation hybride de la PCA et de la LFA. Aucun résultat sur la performance de la reconnaissance basée sur la LFA n'a été rapporté. La LFA est prétendue être utilisée dans le système commercial "FaceIt" de Visionics (Tableau 3.2).

Une méthode flexible basée sur un modèle d'apparence pour la reconnaissance automatique de visage a été présentée dans [Lanitis et autres 1995]. Pour identifier un visage, la forme et l'information du niveau de gris sont modélées et utilisées. Le modèle de forme est un ASM; ce sont les modèles statistiques des formes des objets qui se déforment itérativement pour s'adapter à un exemple de la forme dans une nouvelle image. Le modèle statistique de forme est appris sur des images exemples en utilisant la PCA, où les variables sont les coordonnées des points du modèle de la forme. Pour le but de la classification, les variations de forme dues à la variation inter classes sont séparées de ceux dues aux variations intra classes (telles que de petites variations d'orientation 3D et d'expression faciale) en utilisant l'analyse discriminante. Basé sur le modèle de la forme moyenne, un modèle global de niveau de gris à forme libre peut être construit, encore en utilisant PCA. Pour améliorer la robustesse du système contre des changements d'apparence locale telle que les occlusions, des modèles locaux de niveau de gris sont

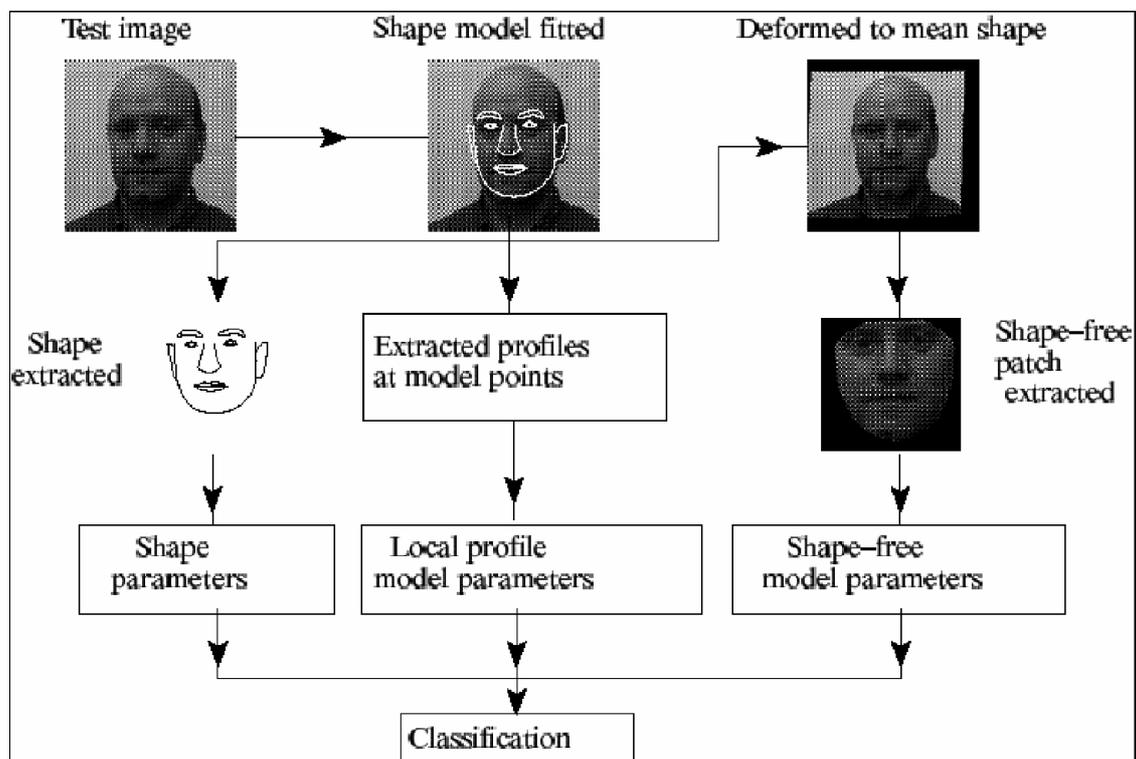
également établis sur les points du modèle de forme. Des profils locaux simples perpendiculaires à la frontière de la forme sont employés. Finalement, pour une image d'entrée, chacun des trois types d'information, y compris les paramètres extraits de forme, les paramètres de forme libres de l'image et les profils locaux, est utilisé pour calculer une distance de Mahalanobis pour la classification comme illustré sur la figure 3.13. Basé sur l'apprentissage de 10 images et le test de 13 images pour chacun des 30 individus, le taux de classification était de 92% pour les 10 images normales de test et de 48% pour les trois images difficiles.



**Figure 3.12** Les noyaux  $K(x_i, y)$  de la LFA à différentes grilles  $x_i$  (figure prise de [Penev et Atick 1996]).

La dernière méthode [Huang et autres 2003] que nous passerons en revue dans cette catégorie est basée sur des avancés récentes dans la détection/reconnaissance basée sur les composantes [Heisele et autres 2001] et sur les modèles morphables en 3D [Blanz et Vetter 1999]. L'idée de base des méthodes basées sur les composantes [Heisele et autres 2001] est de décomposer un visage en un ensemble de composantes faciales telles que la bouche et les yeux qui sont interconnectées par un modèle géométrique flexible (notez comment cette méthode est semblable au système EBGM [Okada et autres 1998 ; Wiskott et autres 1997] sauf que des composantes de niveau de gris sont utilisées au lieu des ondelettes de Gabor). La motivation pour l'usage des composantes est que les changements de pose de la tête mènent principalement aux changements des positions des composantes faciales qui pourraient être expliqués par la flexibilité du modèle géométrique. Cependant, un inconvénient majeur du système est qu'il a besoin d'un grand nombre d'images d'apprentissage prises de différents points de vue et sous différentes conditions d'éclairage. Pour surmonter ce problème, le modèle morphable du visage en 3D [Blanz et Vetter 1999] est appliqué pour générer des images synthétiques arbitraires à des poses et des illuminations variables. Seulement trois images faciales (frontale, de profil et de semi profil) d'une personne sont nécessaires pour calculer le modèle du visage

en 3D. Une fois que le modèle en 3D est construit, des images synthétiques de taille  $58 \times 58$  sont produites pour l'apprentissage du détecteur et du classificateur. Spécifiquement, les visages ont subi des rotations hors du plan de  $0^\circ$  à  $34^\circ$  avec incréments de  $2^\circ$  et sont interprétés avec deux modèles d'illumination (le premier modèle se compose d'une lumière ambiante seule et le second inclut une lumière ambiante et une source de point lumineux tournante) à chaque pose. Quatorze composantes faciales ont été utilisées pour la détection de visage, mais seulement neuf composantes qui n'étaient pas fortement recouvertes et qui contenaient les structures du niveau de gris ont été utilisées pour la classification. En plus, la région de visage a été ajoutée aux neuf composantes pour former un vecteur simple de traits (une méthode hybride), qui, plus tard, a été appris par un classificateur SVM [Vapnik 1995]. L'apprentissage sur trois images et le test sur 200 images par individu ont mené aux taux suivants de reconnaissance sur un ensemble de six individus: 90% pour la méthode hybride et approximativement 10% pour la méthode globale qui a utilisé seulement la région de visage; le taux des reconnaissances fausses positives était de 10%.

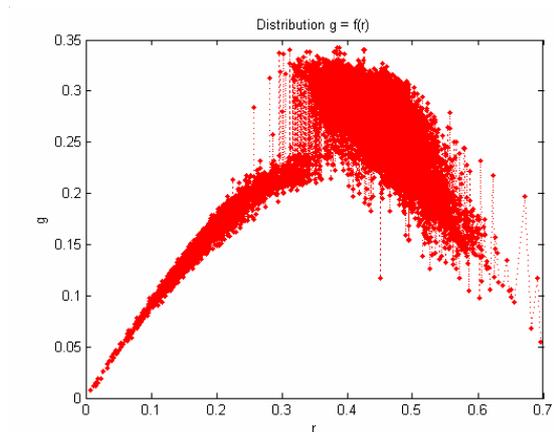


**Figure 3.13** Le schéma de la reconnaissance de visage base sur les modèles flexibles d'apparence ([Lanitis et autres 1995]).

**3.3. Conclusion:**

Dans ce chapitre, on s'est intéressé à la reconnaissance de visage, on a énuméré quelques unes des nombreuses applications de la reconnaissance de visage par machines et des systèmes réelles existant dans l'industrie. On s'est limité à la reconnaissance de visage à partir d'images statiques seulement. Enfin, on a présenté la catégorisation des techniques utilisées pour la reconnaissance, la première catégorie est celle des approches holistiques (PCA, LDA, ...), la deuxième catégorie englobe les méthodes d'assortiment (ou structurelles) basées sur les traits tandis que la troisième catégorie est celle des approches hybrides.

# Chapitre 4: Travail et Contribution



#### **4.1. Introduction:**

Dans ce chapitre, on va exposer avec plus de détails les méthodes choisies pour l'implémentation de la détection et de la reconnaissance de visage. En ce qui concerne la détection de visage, on a proposé deux nouvelles méthodes, la première se base sur la recherche de certains traits caractéristiques du visage humain (les yeux), la deuxième utilise l'information de couleur des pixels de la peau pour isoler les régions susceptibles de contenir des visages humains. Pour la reconnaissance de visage, on a implémenté la méthode des visages propres (l'analyse en composantes principales, PCA) ainsi qu'une autre méthode qui est la méthode des visages propres des images réduites.

#### **4.2. Détection de visage:**

Dans cette section, on va présenter les travaux ou implémentations faites pour la détection de visage, on a proposé deux implémentations pour la détection de visage:

1. La première implémentation utilise en premier lieu la méthode de [Wu et Zhou 2003] pour trouver les candidats de visage moyennant les paires de yeux possibles, ensuite elle vérifie la présence du visage parmi ces candidats de visage en utilisant une méthode qu'on appellera "*Assortiment des projections de la PCA (PCA-projection matching)*".
2. Dans la deuxième méthode, on développe en premier lieu un schéma de segmentation de couleurs dans l'espace des couleurs normalisées rgb, qui nous donne les régions ayant une couleur similaire à celle de la peau, ensuite la vérification de visage se fait en utilisant l'information sur la géométrie du visage et des yeux combinée avec la méthode de [Wu et Zhou 2003].

##### **4.2.1. Première implémentation:**

La première étape de cette implémentation consiste à détecter les paires possibles de yeux de différentes tailles et dans différentes positions dans l'image à niveaux de gris présentée à l'entrée du système, ces paires de yeux ou segments analogues aux yeux (de taille prédéterminée) sont découverts en trouvant les régions qui sont à peu près de la même taille voulue et qui sont plus sombres que leur entourage. Ensuite chaque paire de ces segments analogues aux yeux (*eye-analogue segments*) est supposée constituer des yeux dans un visage et est combinée en un visage si son placement est consistant avec les caractéristiques anthropologiques des yeux humains.

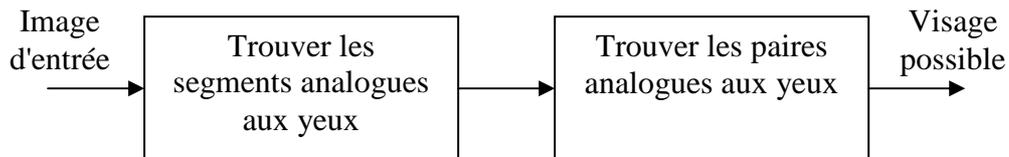
La deuxième partie de cette implémentation consiste à vérifier la présence de visage supposé par la paire de segments analogues aux yeux qui a vérifié les caractéristiques anthropologiques des yeux humains, elle consiste à projeter la partie de l'image qui contient les yeux dans l'espace de visage (face space) de la PCA et de décider ensuite de la présence ou non d'un visage à cet endroit en comparant cette projection à un seuil établi au préalable.

##### **4.2.1.1. Première étape: la détection des yeux et des candidats de visage:**

Cette étape s'intéresse au développement d'un sélecteur de candidats de visage qui localise les formes de visage possibles en combinant deux segments analogues aux yeux. Au début, tous les segments analogues aux yeux sont découverts en trouvant les régions qui sont de même taille qu'une paire de yeux réelle et qui sont

plus sombres que leurs voisines. Si les positions d'une paire de ces segments est conforme aux relations géométriques des yeux d'un visage humain, le sélecteur de candidats de visage conclut qu'une forme de visage peut exister.

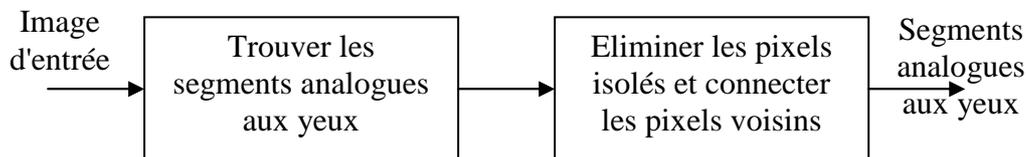
Le bloc diagramme du sélecteur des candidats de visage est montré dans la figure 4.1.



**Figure 4.1** Bloc diagramme du sélecteur des candidats de visage.

#### **4.2.1.1.1. La recherche des segments analogues aux yeux:**

Les yeux sont plus sombres que les autres régions du visage, de sorte que les segments analogues aux yeux sont trouvés en cherchant de petites régions dans l'image d'entrée qui sont si grands comme des yeux et plus sombres que les régions voisines. Le bloc diagramme de la recherche des segments analogues aux yeux est présenté dans la figure 4.2.



**Figure 4.2** Bloc diagramme de la recherche des segments analogues aux yeux.

Considérons une image à niveaux de gris  $P(x, y)$  de taille  $N_1 \times N_2$ , où  $x \in [1, N_1]$ ,  $y \in [1, N_2]$ ,  $P(x, y) \in [0, 1]$ , dans laquelle  $x$  est l'index des lignes et  $y$  l'index des colonnes. Soit  $avg(P, x, y, h, w)$  l'intensité moyenne de la région d'image dont le coin supérieur gauche a pour coordonnées  $(x, y)$  et dont la taille est de  $h \times w$  pixels, c'est-à-dire:

$$avg(P, x, y, h, w) = \frac{\sum_{i=x}^{x+h-1} \sum_{j=y}^{y+w-1} P(i, j)}{hxw} \quad (4.1)$$

L'intensité de chaque pixel dans l'image d'entrée est comparée à l'intensité moyenne de ses huit régions voisines (voir figure 4.3). Heuristiquement, si un pixel est plus sombre que la grande partie de ses voisins, il est marqué comme *un pixel analogue aux yeux (eye-analogue pixel)*. Plus précisément, si l'on cherche des yeux de

taille  $h_e \times w_e$  dans l'image d'entrée, le pixel  $(x, y)$  est préliminairement marqué comme un pixel analogue aux yeux si et seulement si six ou plus de ces contraintes sont satisfaites:

$$P(x, y) < 0.9 * avg(P, x - \lfloor h_e / 2 \rfloor, y - \lfloor w_e / 2 \rfloor, \lfloor h_e / 2 \rfloor, \lfloor w_e / 2 \rfloor) \quad (4.2)$$

$$P(x, y) < 0.9 * avg(P, x - \lfloor h_e / 2 \rfloor, y, \lfloor h_e / 2 \rfloor, 1) \quad (4.3)$$

$$P(x, y) < 0.9 * avg(P, x - \lfloor h_e / 2 \rfloor, y + 1, \lfloor h_e / 2 \rfloor, \lfloor w_e / 2 \rfloor) \quad (4.4)$$

$$P(x, y) < 0.9 * avg(P, x, y - \lfloor w_e / 2 \rfloor, 1, \lfloor w_e / 2 \rfloor) \quad (4.5)$$

$$P(x, y) < 0.9 * avg(P, x, y + 1, 1, \lfloor w_e / 2 \rfloor) \quad (4.6)$$

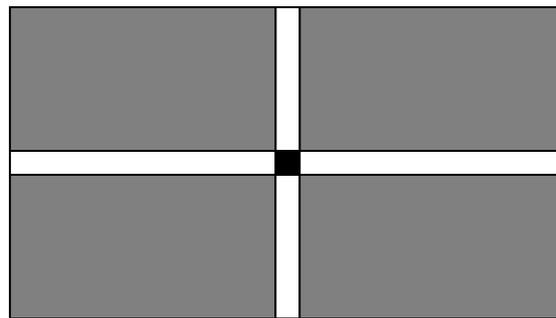
$$P(x, y) < 0.9 * avg(P, x + 1, y - \lfloor w_e / 2 \rfloor, \lfloor h_e / 2 \rfloor, \lfloor w_e / 2 \rfloor) \quad (4.7)$$

$$P(x, y) < 0.9 * avg(P, x + 1, y, \lfloor h_e / 2 \rfloor, 1) \quad (4.8)$$

$$P(x, y) < 0.9 * avg(P, x + 1, y + 1, \lfloor h_e / 2 \rfloor, \lfloor w_e / 2 \rfloor) \quad (4.9)$$

Dans lesquelles  $\lfloor \cdot \rfloor$  est la fonction du plus grand entier (*greatest integer function or the floor function*).

A cause des différentes directions d'illuminations qui peuvent se présenter dans les images de visage, on ne peut exiger qu'un pixel analogue aux yeux soit plus sombre que tous ses huit voisins, six est compromis raisonnable [Wu et Zhou 2003].

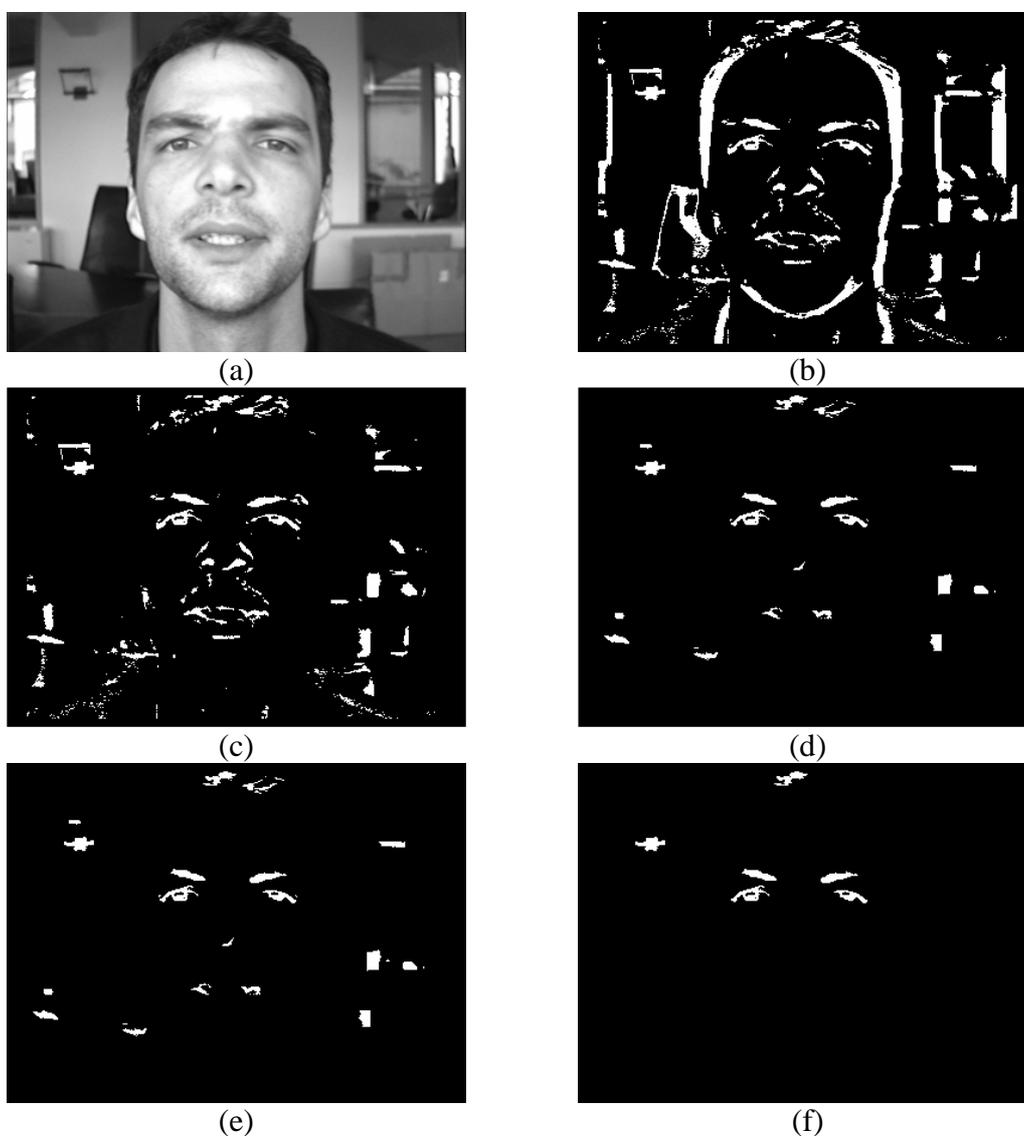


**Figure 4.3** Un pixel (le carré noir) et ses huit régions voisines.

On désigne par  $P'$  l'image formée par les pixels analogues aux yeux à partir de l'image  $P$ .  $P'(x, y)$  est égal à 1 si  $P(x, y)$  est un pixel analogue aux yeux et 0 autrement.

A présent, on présente un exemple pour illustrer le processus de sélection des candidats de visage. La figure 4.4(a) est l'image originale (cette image est tirée de la base de données BioID Face Database, <http://www.bioid.com/research/index.htm>, [Jesorsky et autres 2001]), les pixels analogues aux yeux sont marqués sur la figure 4.4(b) comme des pixels blancs.

Si un groupe de pixels analogues aux yeux forme une région de l'image qui est très large, autrement dit, dont le rectangle qui la contient a une hauteur supérieure à  $2 * h_e$  ou a une largeur supérieure à  $2 * w_e$ , ce groupe n'est pas susceptible de représenter un œil, donc de tels pixels connectés sont éliminés. Les pixels analogues aux yeux restants après cette opération sont montrés sur la figure 4.4(c).



**Figure 4.4** Un exemple du processus de recherche des segments analogues aux yeux.

Ensuite, les pixels isolés sont éliminés. Presque tous les pixels qui se situent entre la paupière supérieure et la paupière inférieure dans une image de visage sont des pixels analogues aux yeux (voir figure 4.4(a)). Si dans le voisinage d'un pixel analogue aux yeux, il n'existe qu'un tout petit nombre de pixels analogues aux yeux, ce pixel doit être éliminé. Plus précisément, un pixel analogue aux yeux  $P'(x, y)$  est éliminé si:

$$avg(P', x - \lfloor h_e / 2 \rfloor, y - \lfloor w_e / 2 \rfloor, h_e, w_e) < 0.2 \quad (4.10)$$

Après l'élimination de tels pixels, les pixels analogues aux yeux sont montrés dans la figure 4.4(d). Dans l'équation 4.10, le seuil 0.2 est très petit parce que cette opération cherche seulement à éliminer les pixels dans les régions extrêmement isolés.

Dans la région des yeux, les paupières (supérieure et inférieure), la pupille et les creux des orbites des yeux sont relativement plus sombres que les autres régions de l'œil. Donc il existe encore peu de pixels avoisinant ces parties qui n'ont pas encore été marqués comme des pixels analogues aux yeux après les opérations précédentes. Par la suite, un pixel est marqué comme un pixel analogue aux yeux s'ils existent suffisamment de pixels analogues aux yeux dans son voisinage. Plus précisément, un pixel  $P(x, y)$  est marqué comme un pixel analogue aux yeux si:

$$avg(P', x - \lfloor h_e / 4 \rfloor, y - \lfloor w_e / 4 \rfloor, \lfloor h_e / 2 \rfloor, \lfloor w_e / 2 \rfloor) > 0.35 \quad (4.11)$$

Le résultat de cette opération est montré dans la figure 4.4(e). Comme dans l'équation 4.10, le seuil 0.35 n'est pas strict lui aussi.

La dernière opération de cette étape consiste à éliminer les régions formées par des pixels connectés analogues aux yeux et qui ne sont pas semblables aux yeux: les régions qui sont trop larges, trop petites et les régions dont l'orientation est presque verticale. Si le rectangle englobant une région de l'image est de taille  $hxw$ , il sera éliminé si l'une quelconque de ces conditions est satisfaite:

$$h / h_e < 0.5 \quad \text{ou} \quad h / h_e > 2.0 \quad (4.12)$$

$$w / w_e < 0.5 \quad \text{ou} \quad w / w_e > 2.0 \quad (4.13)$$

$$h / w > 0.8 \quad (4.14)$$

Les conditions 4.12 et 4.13 excluent les régions extrêmement petites ou extrêmement larges. Notons que les paramètres 0.5 et 2.0 sont assez relaxés parce qu'on ne veut perdre aucune forme de visage, les candidats de visages faux peuvent être éliminés par la deuxième étape de vérification des visages. La valeur moyenne de  $h/w$  pour les yeux humains est approximativement égale à 0.5. L'équation 4.14 exclut les régions ayant un rapport  $h/w$  très large. Ensuite, chaque région restante dans l'image, constituée de pixels connectés analogues aux yeux est un segment analogue aux yeux. Le résultat final est montré dans la figure 4.4(f).

#### **4.2.1.1.2. La détermination des candidats de visage:**

Une région de l'image d'entrée est marquée comme un possible visage si la distance de deux segments analogues aux yeux est comme celle de deux yeux dans un visage humain. Soient  $(x_i, y_i)$  et  $(x_j, y_j)$  les centroïdes de deux segments analogues aux yeux. Un candidat de visage peut exister si toutes les contraintes suivantes sont satisfaites:

$$d_{ij} < 2.5w_e \quad (4.15)$$

$$d_{ij} > 1.5w_e \quad (4.16)$$

$$|x_i - x_j| < h_e \quad (4.17)$$

Dans lesquelles:  $d_{ij}$  est la distance entre les deux centroïdes, c'est-à-dire:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (4.18)$$

Les paramètres des équations 4.15 et 4.16 sont choisis en accord avec le fait qu'en moyenne:  $d_{ij} = 2w_e$ .

#### **4.2.1.2. Deuxième étape: la vérification des candidats de visage par projection dans l'espace de visage de la PCA:**

Après la découverte des candidats possibles de visage dans l'image d'entrée, vient la phase de vérification de ces candidats, pour cela, on dispose de l'espace de visage généré par l'analyse en composantes principales (PCA). La technique de la PCA est généralement utilisée pour la reconnaissance de visage en cherchant dans l'espace de visage, parmi les formes des visages (faces' patterns) utilisées dans l'apprentissage, la forme (pattern) qui constitue le voisin le plus proche de la forme obtenue en projetant l'image de test (ou d'entrée) dans l'espace de visage, tandis que pour la vérification, on a défini un seuil qui nous permet de classer les régions contenant les candidats possibles de visage en visages ou en non visages. Pour cela, on calcule la projection, dans l'espace de visage, du candidat possible de visage, si la distance entre cette projection et la projection de l'image moyenne de l'ensemble d'apprentissage est inférieure au seuil prédéfini, le candidat de visage présente vraiment un visage, si cette distance est supérieure au seuil, ce candidat de visage ne contient pas de visage humain et les segments analogues aux yeux de cette région ainsi que ce candidat de visage sont ignorés.

##### **4.2.1.2.1. La création de l'espace de visage:**

Pour la création de l'espace de visage, on a utilisé un ensemble d'apprentissage constitué de 100 images extraites de la base de données ORL Face Database [ORL Database], après que ces images aient subi une égalisation de l'histogramme dans le but d'éliminer les effets de la variation de luminosité, le processus d'apprentissage est exécuté pour obtenir l'espace propre, les visages (vecteurs) propres, l'image moyenne et les projections des images d'apprentissage dans l'espace de visage.

Ensuite vient la phase de détermination du seuil de détection, chaque image  $\Gamma_i$  de l'ensemble d'apprentissage est projetée dans l'espace de visage pour obtenir la forme de visage  $\Omega_i^T = [\omega_1, \omega_2, \dots, \omega_M]$  correspondante par l'équation:

$$\omega_k = u_k^T (\Gamma_i - \Psi) \quad k \in [1, m] \quad (4.19)$$

Le seuil de détection est choisi comme la distance maximale entre les formes des visages d'apprentissage et la forme de l'image moyenne  $\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i$ , la forme de l'image moyenne  $\Psi$  est égale au vecteur nul dans l'espace de visage du fait qu'avant la projection, l'image moyenne est soustraite de chaque image d'apprentissage, ce qui fait que cette distance se résume au maximum parmi les normes des vecteurs  $\Omega_i^T = [\omega_1, \omega_2, \dots, \omega_M]$   $i \in [1, 100]$  qui représentent les projections des visages d'apprentissage dans l'espace propre, en d'autres mots, le seuil de détection  $\delta$  est obtenu via la formule:

$$\delta = \max(\|\Omega_j\|) \quad \forall j \in [1, 100] \quad (4.20)$$

Pour notre implémentation, on a trouvé:

$$\delta = 2.79e+007 \quad (4.21)$$

Donc, pour chaque candidat possible de visage représenté par la région de l'image  $\Gamma_x$ ,  $\Gamma_x$  est tout d'abord normalisée à une taille égale à celle des images d'apprentissage, ensuite elle subit une égalisation de l'histogramme et sera par la suite projetée dans l'espace de visage pour obtenir la forme  $\Omega_x = [\omega_1, \omega_2, \dots, \omega_M]$ , si on aura:

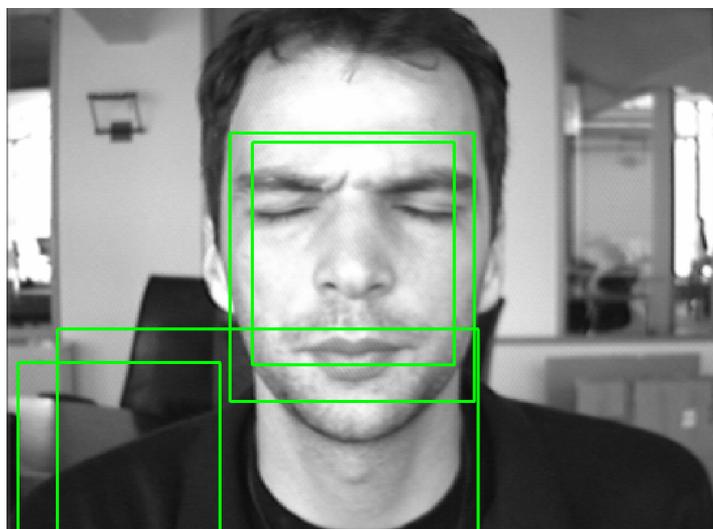
$$\|\Omega_x\| < \delta \quad (4.22)$$

Cette région est marquée comme un visage, sinon, cette région est ignorée (ne contient pas de visage).

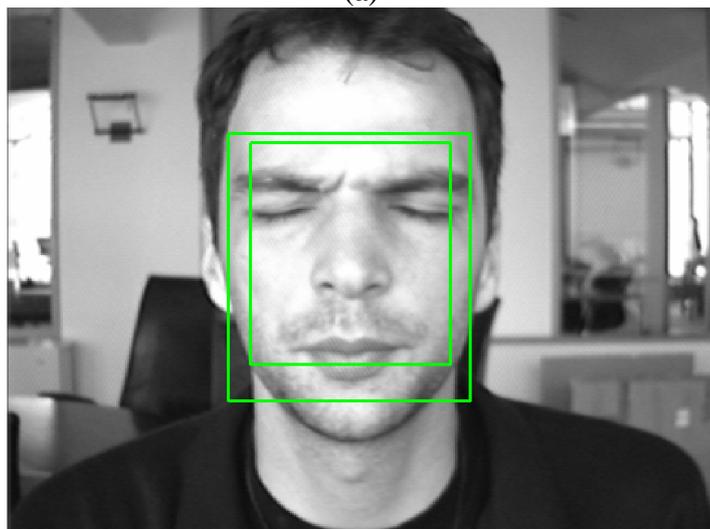
Dans le but de détecter des visages de différentes tailles dans l'image d'entrée deux procédures sont possibles:

- Soit que la taille de l'image d'entrée est réduite à chaque itération d'un facteur (dans notre implémentation, on a pris un facteur de 1.2) tandis que la taille des segments analogues aux yeux est maintenue fixe.
- Soit que la taille de l'image d'entrée est maintenue fixe, tandis que la taille des segments analogues aux yeux recherchés est agrandie d'un facteur (1.2 dans notre implémentation) pour chaque itération.

A présent, on peut illustrer la vérification de visage dans la figure 4.5, la figure 4.5(a) montre les candidats de visage possibles (régions encadrés par des rectangles rouges), la figure 4.5(b) montre les candidats de visage qui ont survécus après le processus de vérification.



(a)



(b)

**Figure 4.5** Processus de vérification des candidats de visage.

#### **4.2.2. Deuxième implémentation:**

La première étape de cette implémentation consiste à développer un schéma de segmentation couleur dans l'espace des couleurs normalisées rgb afin d'isoler les régions de couleur de la peau du reste de l'image et à utiliser par la suite quelques contraintes géométriques pour améliorer le résultat de la segmentation, ce schéma de segmentation couleur est basé sur l'observation des distributions des différentes couleurs (r, g, et b) des pixels de différents images de visages et l'approximations des frontières supérieure et inférieure de ces distributions par des polynômes, ces derniers représentent des fonctions de classification ou de discrimination entre les pixels de la peau et ceux non issus de la peau.

La deuxième partie de cette implémentation consiste à vérifier la présence de visages dans les régions segmentées de la peau, on utilise pour cela la technique de la recherche des segments analogues aux yeux présentée dans la première implémentation, l'avantage de la deuxième implémentation est la réduction significative du temps d'exécution par rapport à la première implémentation, parce que cette dernière requiert plusieurs itérations pour pouvoir détecter des visages de différentes tailles dans l'image d'entrée en plus du temps nécessaire pour vérifier les candidats possibles de visage, tandis que dans la deuxième implémentation, la qualité de la segmentation nous donne des régions de visage dont la largeur nous permet de présumer de la taille des yeux à rechercher et donc d'éviter le grand nombre d'itérations requises dans la première implémentation, on utilise en effet seulement cinq itérations dans la deuxième implémentation. De plus, le processus de segmentation est très rapide.

#### **4.2.2.1. Première étape: développement du schéma de segmentation et des contraintes géométriques:**

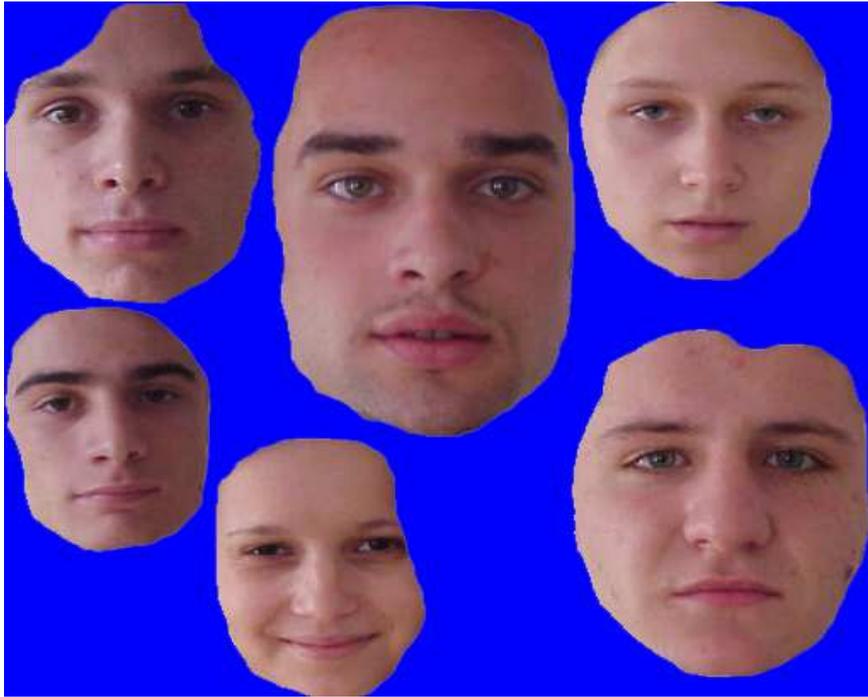
On peut diviser cette étape en deux processus différents:

1. **Développement du schéma de segmentation:** qui consiste à étudier les distributions des couleurs des pixels appartenant à des visages différents dans l'espace des couleurs normalisées rgb et à approximer leurs frontières par des polynômes qui présentent pour des fonctions de discrimination.
2. **Amélioration de la segmentation par des contraintes géométriques:** ce processus consiste à prendre en considération la relations géométriques du visage humain et celles existant entre le visage et les yeux pour éliminer d'avantage de régions ne contenant pas de visages.

##### **4.2.2.1.1. Développement du schéma de segmentation:**

Pour pouvoir observer les distributions des couleurs des pixels de visage dans l'espace de couleurs rgb, on a construit une image qui contient des échantillons de visages rognés pour ne laisser dans l'image totale que les pixels du visage, l'arrière plan est choisi de couleur bleue pure en accordance avec le fait que les visages ne contiennent pas de couleur bleue pure, ce qui n'aura aucune influence sur la distribution des données observées. La figure 4.6 montre l'image ainsi construite et utilisée pour l'apprentissage

Les visages contenus dans cette image d'apprentissage ont été extraits de la base de données "CVL Face Database", *"The face images used in this work have been provided by the Computer Vision Laboratory, University of Ljubljana, Slovenia [CVL FACE DATABASE]"*.



**Figure 4.6** Image utilisée pour le développement du schéma de segmentation.

Ce schéma de segmentation est désigné pour extraire les régions de la peau des visages dans le but de réduire l'étendu de la recherche dans les images d'entrée pour la prochaine étape de la détection de visage. Pour alléger l'influence de la lumière de l'environnement sur l'extraction des pixels de la peau, on adopte la représentation chromatique des couleurs. Pour l'espace des couleurs chromatiques, chaque pixel est représenté par deux valeurs des trois composantes rgb. La conversion à partir de l'espace de couleurs conventionnel RGB vers l'espace des couleurs chromatiques rgb se fait suivant les règles suivantes:

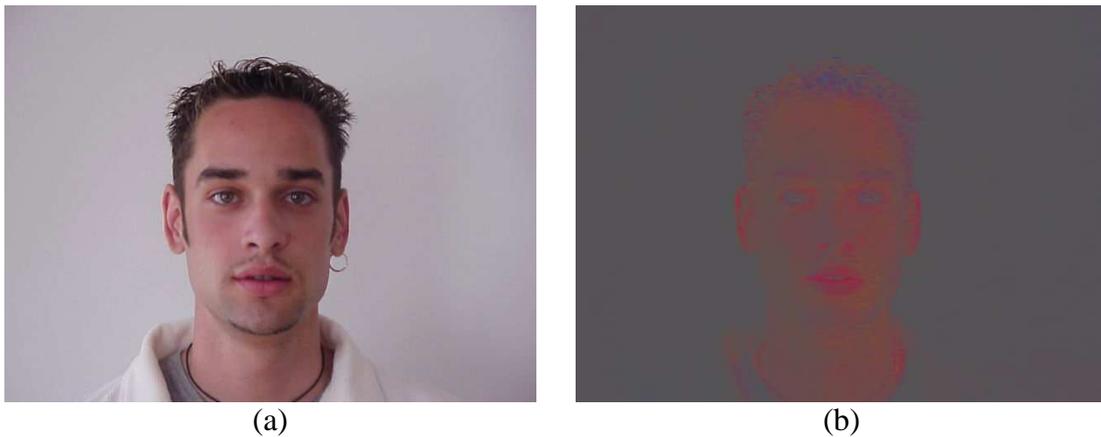
$$r = \frac{R}{R + G + B} \quad (4.23)$$

$$g = \frac{G}{R + G + B} \quad (4.24)$$

$$b = \frac{B}{R + G + B} \quad (4.25)$$

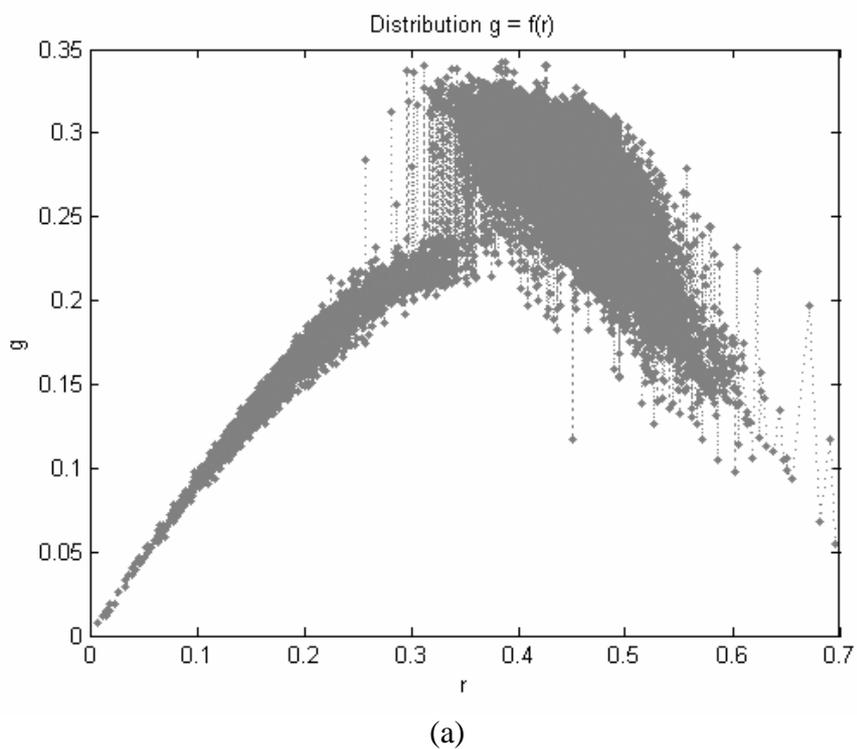
Des équations 4.23, 4.24 et 4.25 on voit que:  $r + g + b = 1$ . Les couleurs normalisées peuvent être effectivement représentées en utilisant seulement deux valeurs, par exemple les valeurs  $r$  et  $g$  puisque la valeur  $b$  peut être obtenue par:  $b = 1 - r - g$ .

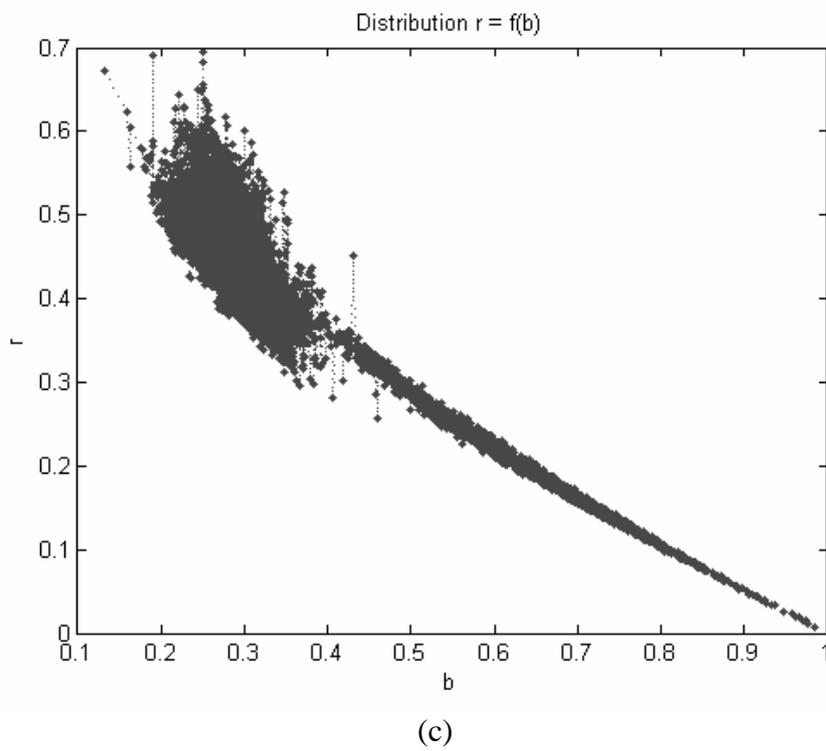
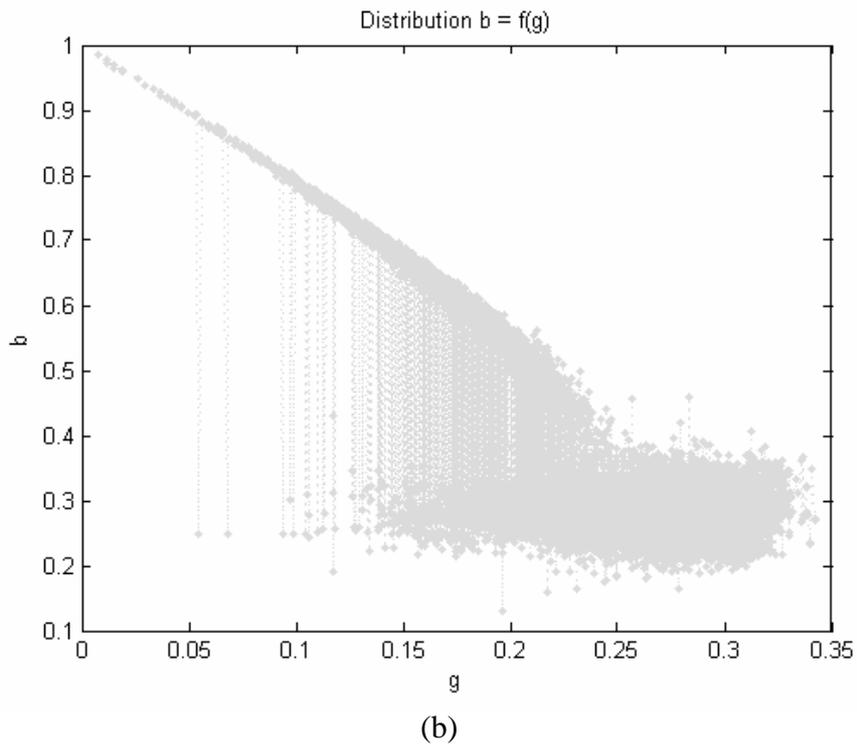
Le résultat de la conversion de l'espace RGB vers l'espace rgb est illustré dans la figure 4.7, la figure 4.7(a) est l'image originale (tirée de CVL Face Database) et la figure 4.7(b) est sa représentation dans l'espace rgb.



**Figure 4.7** Résultat de la conversion  $RGB \rightarrow rgb$ .

A présent, on va présenter dans la figure 4.8 la distribution des couleurs normalisées  $rgb$  des pixels des visages présents dans l'image d'apprentissage (voir figure 4.6).





**Figure 4.8** Distribution des couleurs normalisées rgb des pixels des visages d'apprentissage.

A partir des distributions des couleurs des pixels des visages de la figure 4.8, on a tiré les approximations suivantes des frontières des trois distributions

1. **Distribution  $g=f(r)$ :** approximée par deux polynômes du deuxième degré,  $f_{lower}(r)$  pour la frontière inférieure et  $f_{upper}(r)$  pour la frontière supérieure.

$$f_{lower}(r) = -1.4646r^2 + 1.0788r - 0.0064 \quad (4.26)$$

$$f_{upper}(r) = -3.2387r^2 + 2.3816r - 0.0881 \quad (4.27)$$

2. **Distribution  $b=f(g)$ :** approximée par deux lignes droites (polynômes du premier degré),  $f_{lower}(g)$  pour la frontière inférieure et  $f_{upper}(g)$  pour la frontière supérieure.

$$f_{lower}(g) = 0.1997 \quad (4.28)$$

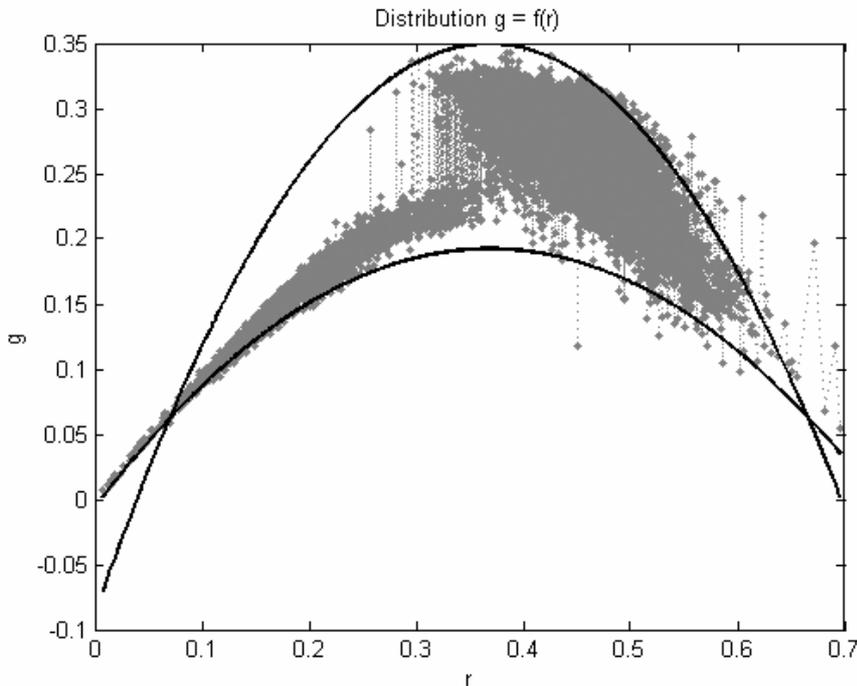
$$f_{upper}(g) = -1.8999g + 1.0023 \quad (4.29)$$

3. **Distribution  $r=f(b)$ :** approximée par deux lignes droites (polynômes du premier degré),  $f_{lower}(b)$  pour la frontière inférieure et  $f_{upper}(b)$  pour la frontière supérieure.

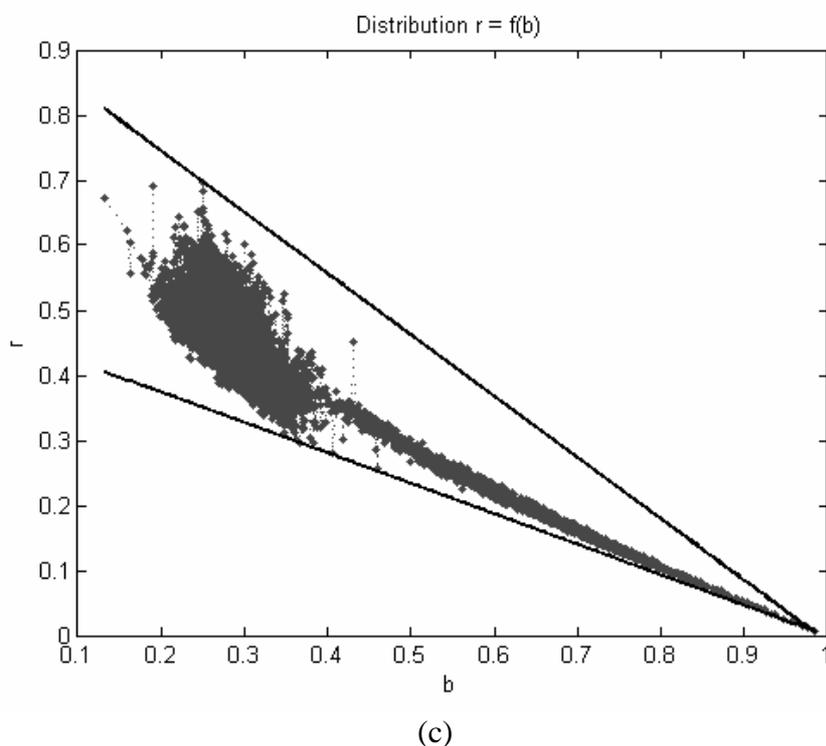
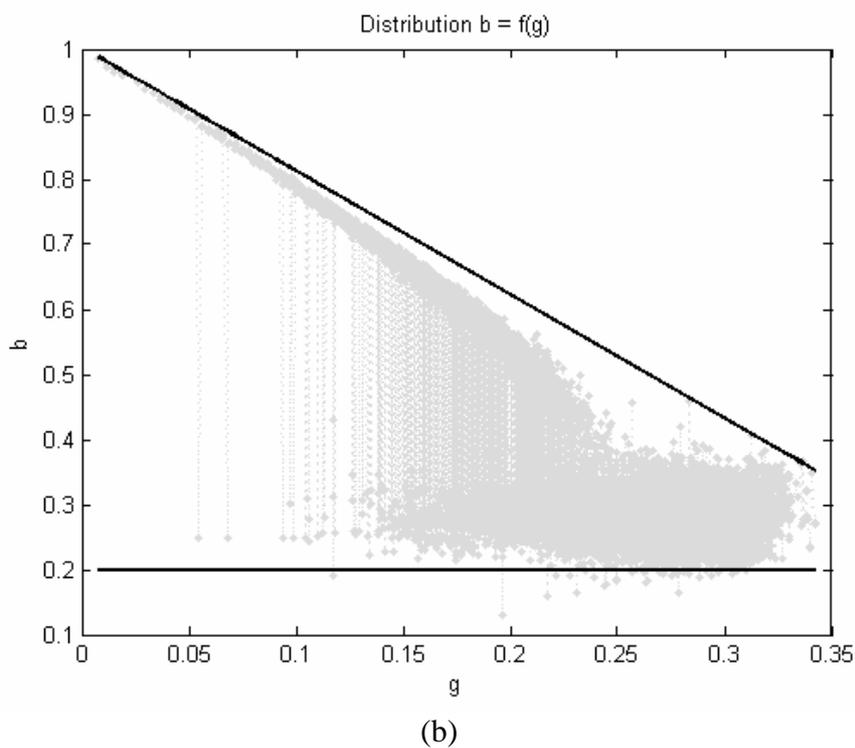
$$f_{lower}(b) = -0.4680b + 0.4676 \quad (4.30)$$

$$f_{upper}(b) = -0.9428b + 0.9327 \quad (4.31)$$

Les trois distributions  $g = f(r)$ ,  $b = f(g)$  et  $r(b)$  ainsi que les approximations de leurs frontières supérieures et inférieures sont illustrées sur la figure 4.9.



(a)



**Figure 4.9** Distribution des couleurs normalisées rgb des pixels des visages d'apprentissage et les approximations de leurs frontières supérieures et inférieures.

En effet, les coefficients de ces polynômes peuvent être estimés facilement par la méthode de la minimisation de l'erreur des moindres carrés (*least mean square error minimization*). Après avoir tracé les frontières de chaque région, on prend (à

titre d'exemple pour la distribution  $g = f(r)$   $n$  points  $(r_1, g_1), (r_2, g_2), \dots, (r_n, g_n)$  de la frontière supérieure, ensuite, on aura l'ensemble des équations linéaires suivantes:

$$\begin{aligned} a_u r_1^2 + b_u r_1 + c_u &= g_1, \\ a_u r_2^2 + b_u r_2 + c_u &= g_2, \\ &\vdots \\ &\vdots \\ &\vdots \\ a_u r_n^2 + b_u r_n + c_u &= g_n, \end{aligned} \quad (4.32)$$

Où  $a_u, b_u$  et  $c_u$  sont les trois coefficients du polynôme quadratique de la frontière supérieure. En réécrivant ces équations sous forme matricielle, on aura:

$$Ru = G \quad (4.33)$$

$$\text{Où: } R = \begin{bmatrix} r_1^2 & r_1 & 1 \\ r_2^2 & r_2 & 1 \\ \vdots & \vdots & \vdots \\ r_n^2 & r_n & 1 \end{bmatrix}, \quad u = \begin{bmatrix} a_u \\ b_u \\ c_u \end{bmatrix} \quad \text{et} \quad G = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix}$$

Donc, le vecteur des coefficients est calculé suivant la formule:

$$u = R^+ G \quad (4.34)$$

Où:  $R^+$  est la matrice pseudo inverse de  $R$ , qui est égale à  $(R^t R)^{-1} R^t$ . Les coefficients du polynôme quadratique de la frontière inférieure peuvent être calculés de la même manière, et ainsi de suite pour les deux autres distributions  $b = f(g)$  et  $r = f(b)$ .

### **Règles de segmentation:**

Pour isoler les pixels de peau des autres pixels de l'image d'entrée, on a défini cinq règles de segmentation:

#### **Règle R1:**

$$g(x, y) > f_{lower}(r(x, y)) \quad \text{et} \quad g(x, y) < f_{upper}(r(x, y)) \quad (4.35)$$

#### **Règle R2:**

$$b(x, y) > f_{lower}(g(x, y)) \quad \text{et} \quad b(x, y) < f_{upper}(g(x, y)) \quad (4.36)$$

#### **Règle R3:**

$$r(x, y) > f_{lower}(b(x, y)) \quad \text{et} \quad r(x, y) < f_{upper}(b(x, y)) \quad (4.37)$$

Les règles R1, R2 et R3 ont été tirées des approximations des frontières des distributions des couleurs de la peau.

**Règle R4:**

$$R(x, y) > G(x, y) > B(x, y) \quad (4.38)$$

Cette règle est dérivée à partir de l'observation que la couleur des pixels de la peau tend vers le rouge et le jaune. Cette observation implique que l'intensité de la composante bleue est toujours la plus petite parmi les trois composantes RGB des pixels de la peau, cette règle permet d'éliminer les pixels dont la couleur tend vers le bleu.

**Règle R5:**

$$R - G \geq 20 \quad (4.39)$$

La règle R5 est définie dans le but d'éliminer les pixel de couleur jaune verte, le paramètre 20 n'est pas stricte et est choisi d'une manière expérimentale.

Donc, pour une image d'entrée  $M(x, y)$ , on définit une image correspondante  $M'(x, y)$  qui est le résultat de la segmentation de l'image  $M(x, y)$  suivant les cinq règles précédentes, c'est-à-dire:

$$M'(x, y) = \begin{cases} 1 & \text{si } M(x, y) \text{ satisfait toutes les règles R1, R2, R3, R4 et R5} \\ 0 & \text{autrement} \end{cases} \quad (4.40)$$

Où  $M'(x, y) = 1$  signifie que le pixel de coordonnées  $(x, y)$  est un pixel de peau et figurera en blanc dans l'image segmentée, si  $M'(x, y) = 0$ , alors ce pixel n'est pas un pixel de peau est figurera en noir dans  $M'(x, y)$ ,

A présent, on peut illustrer, sur la figure 4.10, le processus de segmentation appliqué à une image tirée de la CVL Face Database [CVL FACE DATABASE].



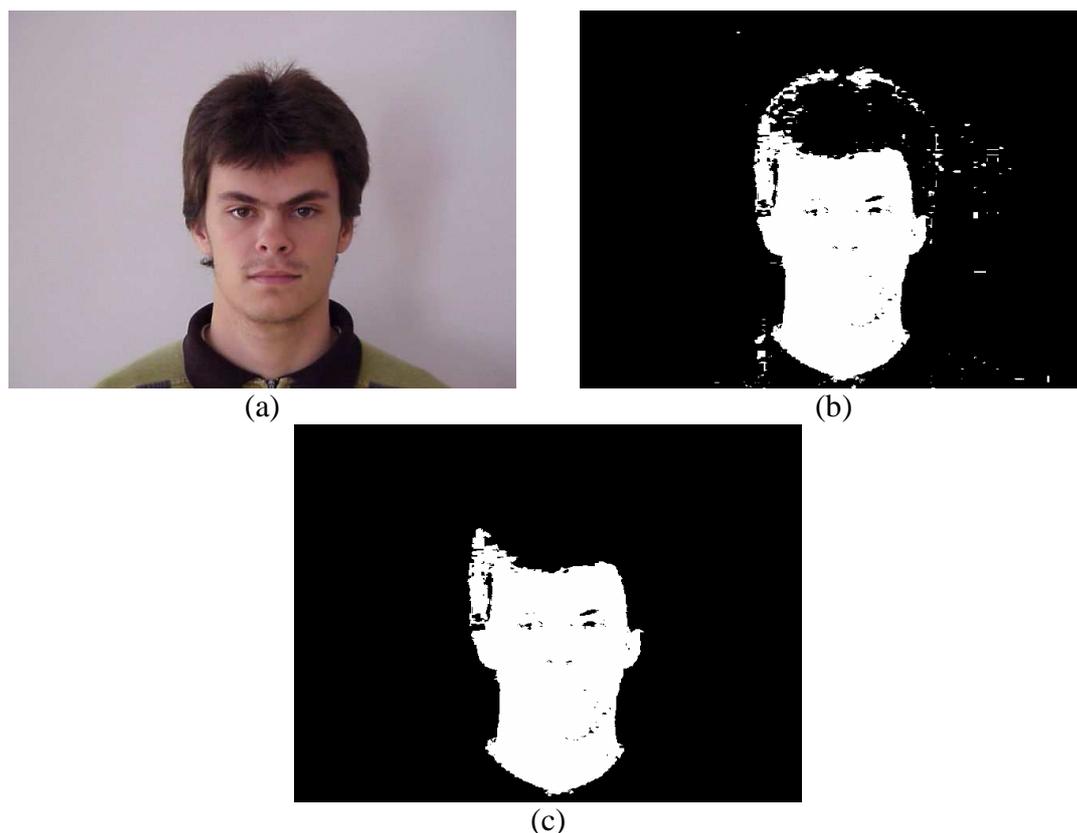
Image originale



Résultat de la segmentation

**Figure 4.10** Le résultat du processus de segmentation appliqué à une image de visage.





**Figure 4.11** Amélioration de la segmentation par des contraintes géométriques.

#### **4.2.2.2. Deuxième étape: La vérification de la présence du visage dans les régions de peau segmentés:**

Le but de cette étape est de vérifier la présence des visages dans les régions segmentées, ceci se fait au moyen de la recherche des segments analogues aux yeux (voir la première étape de la première implémentation) et des candidats possibles des visages dans les régions segmentées de l'image d'entrée, cette recherche est plus rapide que celle dans la première implémentation pour deux raisons essentielles:

1. Le nombre d'itérations est plus petit par rapport à la première implémentation, ceci est dû au fait que la largeur des régions segmentées qui représente la largeur du candidat possible de visage nous permet de présumer de la taille des segments analogues aux yeux qu'on veut découvrir dans l'image.
2. La segmentation permet de réduire le domaine d'application de la recherche des segments analogues aux yeux parce que la taille des régions segmentées est inférieure à celle de l'image d'entrée, en d'autres mots, la segmentation permet d'alléger la recherche des segments analogues aux yeux en réduisant son étendu.

Les détails de la recherche des segments analogues aux yeux et des candidats possibles des visages sont fournis un peu plus haut, dans la section 4.2.1.1.

Pour présumer de la taille des yeux, on a pris, pour des échantillons de visages présents dans la CVL Face Database, des mesures du rapport de la largeur du visage humain à la largeur des yeux, on a trouvé le résultat suivant:

$$\text{moyenne} \left( \frac{\text{largeur}(\text{visage})}{\text{largeur}(\text{oeil})} \right) = 4,6858 \quad (4.41)$$

Donc pour chaque région  $R$  de l'image d'entrée, on définit cinq tailles des yeux à rechercher et qui sont:

$$\left[ \frac{\text{lrg}(R)}{4.6858 * \text{enl}^2}, \frac{\text{lrg}(R)}{4.6858 * \text{enl}}, \frac{\text{lrg}(R)}{4.6858}, \frac{\text{lrg}(R) * \text{enl}}{4.6858}, \frac{\text{lrg}(R) * \text{enl}^2}{4.6858} \right]$$

Où  $\text{lrg}(R)$  signifie largeur de  $R$  et  $\text{enl}$  est un facteur d'agrandissement, dans notre implémentation, on a pris  $\text{enl} = 1.2$ . Le processus de vérification des visages recherche les segments analogues aux yeux et les candidats de visage dans la région

$R$  en commençant par la plus petite taille des yeux  $\frac{\text{lrg}(R)}{4.6858 * \text{enl}^2}$ , ce processus s'arrête lorsqu'un candidat de visage est détecté.

La figure 4.12 montre le résultat final de la détection de visage par cette méthode.



**Figure 4.12** Résultat de la détection de visage.

### **4.3. Reconnaissance de visage:**

Pour la reconnaissance de visage, on a implémenté deux techniques, la première est l'analyse en composantes principales (PCA) avec une comparaison des distances, la deuxième est la méthode des visages propres des images réduites (RIE).

#### **4.3.1. L'analyse en composantes principales (Principal component Analysis, PCA):**

Le succès de la reconnaissance de visage dépend des éléments utilisés pour représenter les images dans le but d'une prochaine classification. La forme représentative d'une image doit être composée par les éléments les plus marquants ou dominants de cette image, permettant de réduire la quantité des données utilisées pendant la classification et d'accentuer les différences entre elles pour que ces formes constituent un classificateur puissant. De plus, ces éléments qu'on appelle aspects (ou attributs) du stimulus du visage, peuvent ou ne pas être directement reliés à notre notion intuitive de traits du visage tels que le nez, les yeux, les lèvres et les cheveux.

Dans le langage de la théorie de l'information, on veut extraire l'information significative (pertinente) d'une image de visage, l'encoder le plus efficacement possible et comparer cet encodage avec une base de données de modèles similairement encodés. Une approche simple pour l'extraction de l'information contenue dans une image de visage est de capturer en quelque sorte la variation dans une collection d'images de visage, indépendamment de tout jugement sur les attributs et utiliser cette information pour encoder et comparer les images de visage.

En termes mathématiques, on cherche à trouver les composantes principales de la distribution des visages, ou la matrice de covariance de l'ensemble des images de visage. Ces vecteurs propres peuvent être considérés comme un ensemble de dispositifs qui, ensemble, caractérisent la variation entre les images de visage. Ces valeurs propres peuvent être affichés sous forme de "visages fantômes (ghostly faces)" qui sont appelés les visages propres (eigenfaces).

Chaque image dans l'ensemble d'apprentissage peut être représentée sous forme de combinaison linéaire de ses visages propres. Le nombre des visages propres possibles est égal au nombre des images de visage présentes dans l'ensemble d'apprentissage. Cependant, ces visages peuvent être approximés en utilisant seulement les "meilleurs visages propres (best eigenfaces)", c'est-à-dire ceux qui correspondent aux valeurs propres les plus grandes, et comptent donc pour la majeure partie de variance dans cet ensemble d'images de visage.

Ces meilleurs visages propres forment la base d'un sous espace - dimensionnel qui est appelé l'espace de visage (face space). Comme les sinusoides de différentes fréquences et de différentes phases sont les fonctions de base d'une décomposition de Fourier (et sont du fait les fonctions propres (eigenfunctions) des systèmes linéaires), les visages propres sont les vecteurs de base de la décomposition du visage propre.

Les premiers utilisateurs de la PCA furent [Kirby et Sirovich 1990] pour la représentation des images de visage, ils ont affirmé qu'une collection d'images de visage peut être approximativement reconstruite en emmagasinant une petite collection de poids pour chaque image de visage et un petit ensemble d'images standards, ensuite [Turk et Pentland 1991] l'ont exploité pour la reconnaissance de visage en construisant des attributs caractéristiques à partir d'images de visages connus et en reconnaissant les autres visages par comparaison des poids des attributs requis pour reconstruire ces nouveaux visages avec les poids associés aux individus connus.

Le processus de reconnaissance se résume dans les étapes suivantes:

1. Initialisation: acquérir l'ensemble d'apprentissage des images de visage et calculer les visages propres qui définissent l'espace de visage.
2. Lorsqu'une nouvelle image est présentée pour la reconnaissance, calculer un ensemble de poids en se basant sur l'image elle-même et sur l'ensemble des  $M$  visages propres, en projetant l'image d'entrée sur chacun des visages propres.
3. Classifier l'image en se basant sur la distance minimale entre sa projection dans l'espace de visage et la projection des images d'apprentissage.

#### 4.3.1.1. L'algorithme PCA:

Considérons une image de visage  $I(x, y)$  de taille  $N \times N$  comme un tableau à deux dimensions dont les valeurs sont les intensités lumineuses (niveaux de gris) des points (pixels) de l'image, cette image peut être considérée comme un vecteur de dimension  $N^2$  ou comme un point dans un espace de dimension  $N^2$ . Donc, on considère que l'ensemble des images d'apprentissage sont des points dans cet espace de très grande dimension  $N^2$ .

Puisque les images de visage sont similaires dans leur configuration globale, elles ne sont pas distribuées aléatoirement dans cet espace de très grande dimension  $N^2$  qui est appelé *l'espace d'image*, elles peuvent être décrites dans un sous espace de dimension relativement petite.

L'idée principale de la PCA est de trouver les vecteurs qui correspondent le plus à la distribution des images de visage à l'intérieur de l'espace d'image. Ces vecteurs définissent le sous espace des images de visage, qui est appelé l'espace de visage.

Chaque vecteur de dimension  $N^2$  (ou de taille  $N^2 \times 1$ ), décrivant une image de taille  $N \times N$ , est une combinaison linéaire des images des visages d'apprentissage. Puisque ces vecteurs sont les vecteurs propres de la matrice de covariance correspondante aux images des visages d'apprentissage et puisqu'ils sont semblables en apparence à des visages, ils sont appelés visages propres.

On utilise comme ensemble d'apprentissage  $M$  images  $I_1, I_2, I_3, \dots, I_M$ , de taille  $N \times N$ .

La figure 4.13 montre quelques unes des images utilisées dans l'ensemble d'apprentissage, ces images appartiennent à la base de données des images de visage ORL.



**Figure 4.13** Quelques unes des images d'apprentissage prises de la base de données ORL.

L'apprentissage consiste en les étapes suivantes:

- a) Chaque image  $I_i$  ( $\forall i \in [1, M]$ ) est réorganisée comme un vecteur  $\Gamma_i$  de taille  $N^2 \times 1$  en concaténant les lignes successives de l'image  $I_i$ .
- b) Le visage moyen est obtenu par la formule:

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad (4.42)$$

La figure 4.14 donne une idée du visage moyen.



**Figure 4.14** Visage moyen obtenu à partir de l'ensemble des images d'apprentissage.

- c) Le visage moyen est soustrait de chacune des images  $\Gamma_i$ , ce qui donne un nouvel ensemble de vecteurs  $\Phi_i, i = 1, 2, \dots, M$  comme suit:

$$\Phi_i = \Gamma_i - \Psi \quad (4.43)$$

Cet ensemble de vecteurs  $\Phi_i, i = 1, 2, \dots, M$  compose la matrice  $\Lambda = [\Phi_1 \Phi_2 \dots \Phi_M]$  de taille  $N^2 \times M$ .

- d) A présent, on cherche les vecteurs propres et les valeurs propres de la matrice de covariance  $C$  :

$$C = \frac{1}{M} \Lambda \Lambda^T \quad (4.44)$$

La matrice  $C$  est de taille  $N^2 \times N^2$ .

Les vecteurs propres sont les vecteurs ortho normaux utilisés pour former le sous espace de visage et pour décrire, le mieux possible, la distribution des visages dans l'ensemble d'apprentissage. La taille de la matrice  $C$  rend cette étape impossible (espace mémoire et temps d'exécution requis) même pour des images de visage de taille réduite, c'est pourquoi on cherche une approximation de ces vecteurs propres.

- d.1) La matrice de covariance réduite  $L$  de dimension  $M \times M$  est obtenue comme suit:

$$L = \frac{1}{M} \Lambda^T \Lambda \quad (4.45)$$

Avec cette analyse, les calculs sont fortement réduits de l'ordre du nombre de pixels dans l'image de visage ( $N^2$ ) au nombre d'images dans l'ensemble d'apprentissage ( $M$ ). Dans la pratique, l'ensemble d'apprentissage des images de

visage sera relativement petit ( $M \ll N^2$ ) et les calculs peuvent être effectués sur un ordinateur.

d.2) Les vecteurs propres  $v_i$  de taille  $M \times 1$  et les valeurs propres  $\lambda_i$  de la matrice de covariance réduite  $L$  sont alors obtenus via la formule:

$$\lambda_i v_i = L v_i \quad i = 1, 2, \dots, M \quad (4.46)$$

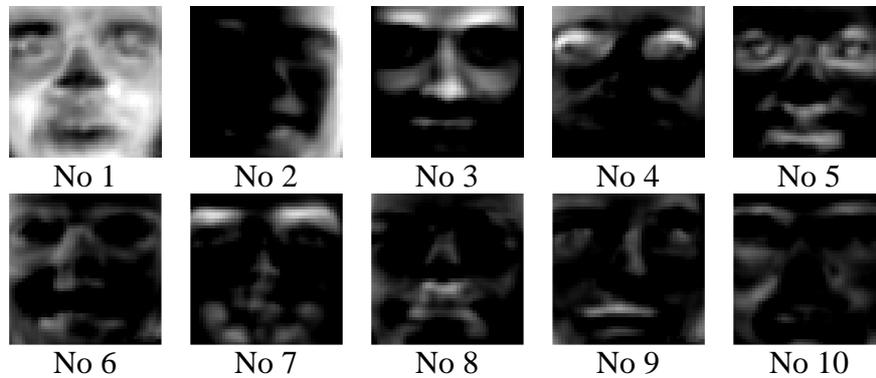
Ces vecteurs propres sont ensuite ordonnés suivant l'ordre décroissant de leurs valeurs propres correspondantes  $\lambda_i \quad i = 1, 2, \dots, M$ , ils sont ensuite concaténés colonne par colonne pour former la matrice  $V$  de taille  $M \times M$ .

d.3) Les vecteurs propres  $u_i$  ( $i = 1, 2, \dots, M$ ) de taille  $N^2 \times 1$  de la matrice  $C$  sont approximés par la formule:

$$U = \Lambda V \quad (4.47)$$

Où  $U = [u_1 u_2 \dots u_M]$  est de taille  $N^2 \times M$ .

La figure 4.15 montre des exemples de visages propres ordonnés en ordre décroissant suivant leurs valeurs propres correspondantes.



**Figure 4.15** Exemples de visages propres.

e) Une forme (projection dans l'espace de visage) de l'image  $i$  ( $\forall i \in [1, M]$ ) de l'ensemble d'apprentissage est le vecteur  $\Omega_i^T = [\omega_1, \omega_2, \dots, \omega_M]$  où:

$$\omega_k = u_k^T (\Gamma_i - \Psi) \quad k = 1, 2, \dots, M \quad (4.48)$$

#### **4.3.1.1.1. Processus de reconnaissance:**

Etant donnée une nouvelle image  $\Gamma_x$ , le processus de reconnaissance cherche à trouver dans l'ensemble d'apprentissage, l'image de visage qui correspond le mieux à notre nouvelle image. Ceci est accompli en transformant l'image d'entrée  $\Gamma_x$  en ses coefficients de visages propres, c'est à dire en la projetant dans l'espace de visage via la formule:

$$\Omega_x = U^T (\Gamma_x - \Psi) \quad (4.49)$$

Ce vecteur de poids  $\Omega_x^T = [\omega_1, \omega_2, \dots, \omega_M]$  décrit la contribution de chaque visage propre dans la représentation de l'image d'entrée  $\Gamma_x$ , en considérant les visages propres comme un ensemble de base pour les images de visage (vecteurs de base de l'espace de visage). Ce vecteur est utilisé pour trouver le visage qui décrit le mieux possible le visage d'entrée parmi les visages utilisés lors de l'apprentissage.

Le visage  $\Gamma_i$  qui ressemble le plus au visage de l'image d'entrée  $\Gamma_x$  est celui qui satisfait la relation:

$$\|\Omega_i - \Omega_x\|^2 = \min \left( \|\Omega_j - \Omega_x\|^2 \right) \quad \forall j \in [1, M] \quad (4.50)$$

#### **4.3.2. La méthode des visages propres des images réduites (Reduced Image Eigenfaces, RIE):**

Cette méthode a été proposée par [Chichizola et autres 2005], de ce qui a été mentionné précédemment dans l'algorithme PCA, on cherche à calculer les vecteurs propres de la matrice de covariance. Vu leur très grande taille, ils sont approximés par le biais des vecteurs propres de la matrice de covariance réduite (voir équation 4.47).

Cette méthode consiste en la transformation des images de visage en des images plus petites dans le but de travailler directement avec la matrice de covariance au lieu d'utiliser son approximation.

L'apprentissage consiste en les étapes suivantes:

- a) Chaque image  $I_i$  ( $\forall i \in [1, M]$ ) est divisée en blocs de taille  $P \times P$  pixels chacun.  $P$  étant le *taux de réduction (reduction level)*, la moyenne de chacun de ces blocs est ensuite calculée et une nouvelle image  $I_i'$  de taille  $D \times D$  est calculée, avec  $D = N / P$ , cette image  $I_i'$  est obtenue en remplaçant chaque bloc dans l'image originale  $I_i$  par sa moyenne.

La figure 4.16 donne une idée de l'opération de réduction pour quelques images de l'ensemble d'apprentissage et pour différents taux de réduction.

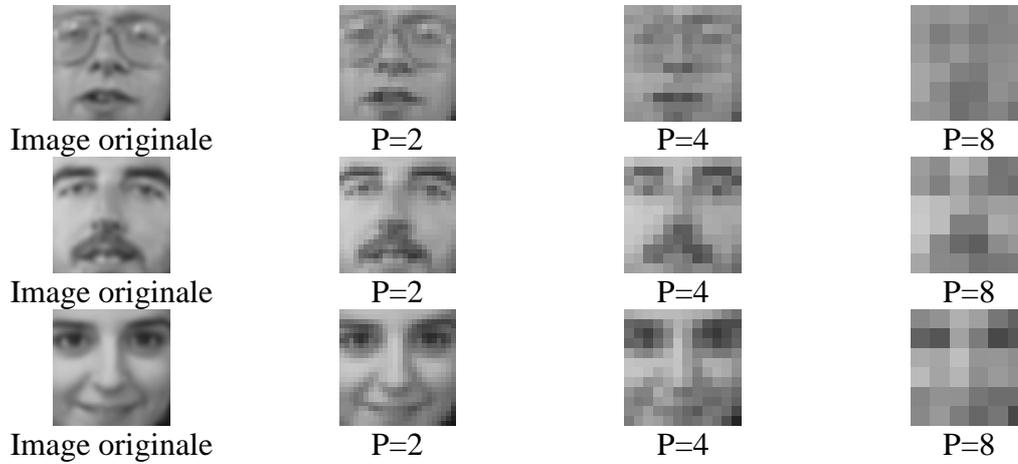
- b) Chaque image  $I_i'$  est réorganisée comme un vecteur colonne  $\Gamma_i$  de taille  $D^2 \times 1$  en concaténant les lignes successives de l'image  $I_i'$ .
- c) Le visage moyen est obtenu par la formule:

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad (4.51)$$

- d) Le visage moyen est soustrait de chacune des images  $\Gamma_i$ , ce qui donne un nouvel ensemble de vecteurs  $\Phi_i, i = 1, 2, \dots, M$  comme suit:

$$\Phi_i = \Gamma_i - \Psi \quad (4.52)$$

Cet ensemble de vecteurs  $\Phi_i, i = 1, 2, \dots, M$  compose la matrice  $\Lambda = [\Phi_1 \Phi_2 \dots \Phi_M]$  de taille  $D^2 \times M$ .



**Figure 4.16** Effet de la réduction sur les images de visage de l'ensemble d'apprentissage.

e) La matrice de covariance est obtenue par la formule:

$$C = \frac{1}{M} \Lambda \Lambda^T \quad (4.53)$$

La matrice  $C$  est de taille  $D^2 \times D^2$ .

f) Les vecteurs propres de  $C$  sont alors calculés, ils sont ensuite ordonnés en ordre décroissant suivant leurs valeurs propres correspondantes et concaténés colonne par colonne pour former la matrice  $U$  où  $U = [u_1 u_2 \dots u_{D^2}]$  est de taille  $D^2 \times D^2$ .

g) Une forme (projection dans l'espace de visage) de l'image  $i$  ( $\forall i \in [1, M]$ ) de l'ensemble d'apprentissage est le vecteur  $\Omega_i^T = [\omega_1, \omega_2, \dots, \omega_{D^2}]$  où:

$$\omega_k = u_k^T (\Gamma_i - \Psi) \quad i = 1, 2, \dots, D^2 \quad (4.54)$$

Le processus de reconnaissance est similaire à celui de l'algorithme PCA classique, étant donnée une nouvelle image réduite  $\Gamma_x$ , sa projection  $\Omega_x^T = [\omega_1, \omega_2, \dots, \omega_{D^2}]$  est calculée tout d'abord par l'équation:

$$\Omega_x = U^T (\Gamma_x - \Psi) \quad (4.55)$$

Ensuite, on cherche le visage  $\Gamma_i$  qui ressemble le plus au visage de l'image d'entrée  $\Gamma_x$  en satisfaisant l'équation:

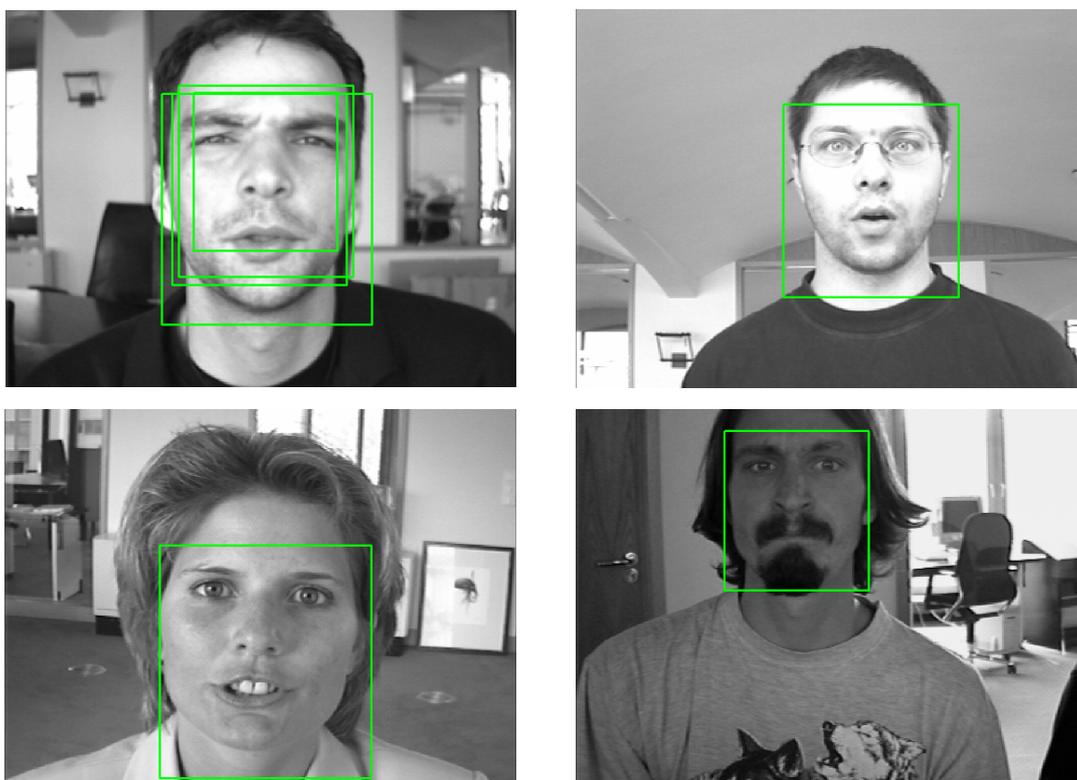
$$\|\Omega_i - \Omega_x\|^2 = \min \left( \|\Omega_j - \Omega_x\|^2 \right) \quad \forall j \in [1, M] \quad (4.56)$$

#### **4.4. Expérimentations et résultats:**

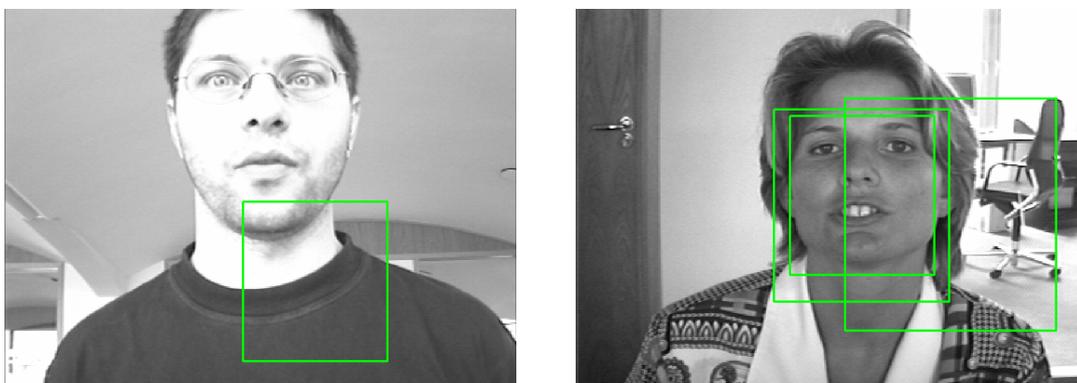
##### **4.4.1 Détection de visage:**

###### **4.4.1.1. Première implémentation:**

Dans les figures suivantes, on va présenter quelques résultats de la détection de visage obtenus avec la première implémentation, en utilisant les images provenant de la base de données BioID Face Database. La figure 4.17 montre des cas où la détection de visage est réussie, la figure 4.18 montre des cas de fausses détections.



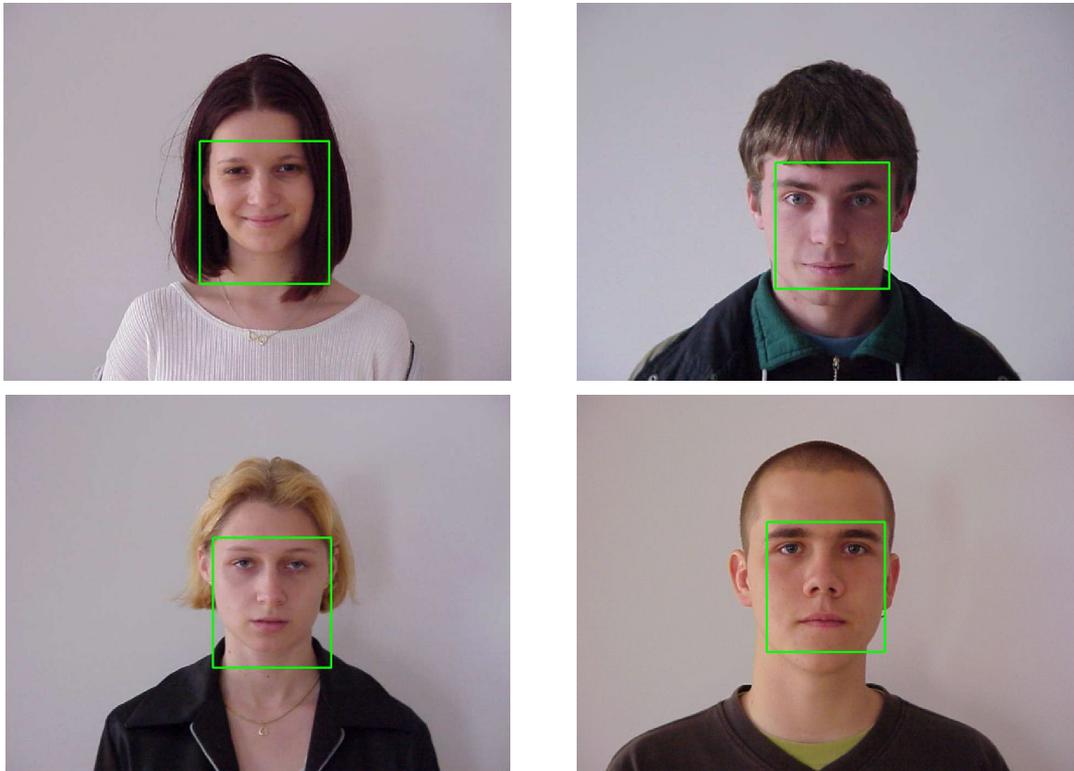
**Figure 4.17** Visages correctement détectés.



**Figure 4.18** Détections fausses de visages.

**4.4.1.2. Deuxième implémentation:**

Les tests de la deuxième implémentation ont été effectués sur des images de visage provenant de la base de données CVL Face Database. La figure 4.19 montre des exemples de détections correctes de visages tandis que la figure 4.20 montre des détections fausses du visage.



**Figure 4.19** Détections correctes de visages.



**Figure 4.20** Détections fausses de visages.

#### **4.4.2. Reconnaissance de visage:**

##### **4.4.2.1. L'analyse en composantes principales (PCA):**

En ce qui concerne l'algorithme classique de la PCA, les tests ont été réalisés sur deux bases de données d'images de visage, à savoir la base de données BioID (*BioID face database*) et la base de données ORL (*ORL face database*).

##### **4.4.2.1.1. Résultats obtenus avec la base de données BioID:**

On a choisi 3 configurations pour les deux ensembles de test et d'apprentissage:

- **1ère configuration:** l'ensemble d'apprentissage comprend 5 individus différents, 1 image de visage par individu, c'est à dire 5 images de visage, l'ensemble de test est constitué de 45 images de visage, c'est à dire 9 images de test pour chacun des 5 individus présents dans l'ensemble d'apprentissage, au total 50 images de visage sont utilisés pour l'apprentissage et le test.
- **2ème configuration:** l'ensemble d'apprentissage comprend 10 individus différents, 1 image de visage par individu, c'est à dire 10 images de visage, l'ensemble de test est constitué de 90 images de visage, c'est à dire 9 images de test pour chacun des 10 individus présents dans l'ensemble d'apprentissage, au total 100 images de visage sont utilisés pour l'apprentissage et le test.
- **3ème configuration:** l'ensemble d'apprentissage comprend 5 individus différents, 5 images de visage par individu, c'est à dire 25 images de visage, l'ensemble de test est constitué de 25 images de visage, c'est à dire 5 images de test pour chacun des 5 individus présents dans l'ensemble d'apprentissage, au total 50 images de visage sont utilisés pour l'apprentissage et le test.
- **4ème configuration:** l'ensemble d'apprentissage comprend 5 individus différents, 9 images de visage par individu, c'est à dire 45 images de visage, l'ensemble de test est constitué de 5 images de visage, c'est à dire 1 image de test pour chacun des 5 individus présents dans l'ensemble d'apprentissage, au total 50 images de visage sont utilisés pour l'apprentissage et le test.

Le tableau suivant résume les différentes configurations utilisées.

Configuration	Ensemble d'apprentissage		Ensemble de test	
	Nombre d'individus	Nombre d'images	Nombre d'individus	Nombre d'images
No 1	5	5	5	45
No 2	10	10	10	90
No 3	5	25	5	25
No 4	5	45	5	5

**Tableau 4.1** Les différentes configurations utilisées avec BioID.

L'intérêt qui réside dans l'utilisation de ces configurations est de pouvoir étudier l'effet de la variation des paramètres suivants sur le taux de reconnaissance de l'algorithme PCA:

- Nombre d'images par individu dans l'ensemble d'apprentissage.

- Nombre d'individus dans les ensembles d'apprentissage et de test.
- Taille des ensembles d'apprentissage et de test.
- Rapport du nombre d'images de l'ensemble de test au nombre d'images de l'ensemble d'apprentissage.

Avant de s'attaquer aux résultats obtenus avec la base de données BioID face database, on présentera dans les figures suivantes quelques unes des images BioID utilisées en test et/ou en apprentissage.



**Figure 4.21** Exemples d'images utilisées dans l'algorithme des visages propres.

D'après les images de la figure 4.21, on s'aperçoit que:

- La luminosité du visage est variable.
- La position du visage est variable.
- La taille du visage est variable.
- L'arrière-plan est variable.
- L'expression faciale est variable.

Ces conditions qui varient d'une image à l'autre constituent le défi principal pour le succès (relatif) d'une technique de reconnaissance du visage.

On présentera dans le tableau suivant les résultats obtenus avec BioID face database, sachant que le taux de reconnaissance est le rapport du nombre de détections correctes sur le nombre des images de l'ensemble de test, c'est-à-dire:

$$\text{Taux de reconnaissance} = \frac{\text{Nombre de détections correctes}}{\text{Nombre d'images de test}} \quad (4.57)$$

Configuration	Nombre de détections correctes	Nombre de fausses détections	Taux de reconnaissance
No 1	19	26	42.22 %
No 2	46	44	51.11 %
No 3	23	2	92 %
No 4	5	0	100 %

**Tableau 4.2** Résultats de la reconnaissance pour 4 configurations différentes avec BioID.

#### **4.4.2.1.2. Résultats obtenus avec la base de données ORL:**

On a choisi 2 configurations pour les deux ensembles de test et d'apprentissage:

- **1ère configuration:** l'ensemble d'apprentissage comprend 20 individus différents, 1 image de visage par individu, c'est à dire 20 images de visage, l'ensemble de test est constitué de 180 images de visage, c'est à dire 9 images de test pour chacun des 20 individus présents dans l'ensemble d'apprentissage, au total 200 images de visage sont utilisés pour l'apprentissage et le test.
- **2ème configuration:** l'ensemble d'apprentissage comprend 20 individus différents, 5 images de visage par individu, c'est à dire 100 images de visage, l'ensemble de test est constitué de 100 images de visage, c'est à dire 5 images de test pour chacun des 20 individus présents dans l'ensemble d'apprentissage, au total 200 images de visage sont utilisés pour l'apprentissage et le test.

Configuration	Ensemble d'apprentissage		Ensemble de test	
	Nombre d'individus	Nombre d'images	Nombre d'individus	Nombre d'images
No 1	20	20	20	180
No 2	20	100	20	100

**Tableau 4.3** Résumé des différentes configurations utilisées avec ORL.

Le tableau suivant montre les résultats obtenus avec ORL face database.

Configuration	Nombre de détections correctes	Nombre de fausses détections	Taux de reconnaissance
No 1	88	92	48.89 %
No 2	79	21	79 %

**Tableau 4.4** Résultats de la reconnaissance pour 2 configurations différentes avec ORL.

#### **4.4.2.1.3. Comparaison de distances:**

Dans le but de comparer les performances et les résultats obtenues en utilisant les différents types de distances, telles que: la distance Euclidienne, la distance de Manhattan, la distance de Chebyshev,... On a utilisé, pour les expérimentations des distances, la deuxième configuration, c'est-à-dire 100 images (20 individus ayant chacun 5 images) pour l'ensemble d'apprentissage ainsi que 100 images (20 individus ayant chacun 5 images) pour l'ensemble de test. Pour une discussion détaillée de ces distances, le lecteur est dirigée vers l'annexe A de cette thèse.

Le tableau 4.5 présente les résultats obtenus avec ces distances.

Distance	Nombre de détections correctes	Nombre de fausses détections	Taux de reconnaissance
Euclidienne	79	21	79 %
Chebyshev	72	28	72 %
Manhattan	80	20	80 %
Minkowski(8)*	76	24	76 %
Angle Based	78	22	78 %
Correlation Coefficient Based	79	21	79 %

\* la distance de Minkowski d'ordre 8.

**Tableau 4.5** Comparaison du taux de reconnaissance de la PCA combinée avec plusieurs types de distances.

En ce qui concerne les distances "Angle Based distance" et "Correlation Coefficient based distance", elles ont été utilisées dans [Perlibakas 2004], qui a comparé 14 distances et leurs dérivées du point de vue des performances de la reconnaissance de visage par la PCA.

#### **4.4.2.2. Résultats obtenus avec la méthode RIE:**

Les ensembles d'apprentissage et de test utilisés avec l'algorithme des visages propres des images réduites sont ceux de la deuxième configuration de la base de données ORL:

- L'ensemble d'apprentissage comprend 20 individus différents, 5 images de visage par individu, c'est à dire 100 images de visage, l'ensemble de test est constitué de 100 images de visage, c'est à dire 5 images de test pour chacun des 20 individus présents dans l'ensemble d'apprentissage, au total 200 images de visage sont utilisés pour l'apprentissage et le test.
- Le taux de réduction choisi pour l'implémentation est:  $P = 4$ .

Le tableau suivant donne les résultats de la technique RIE.

	Nombre de détections correctes	Nombre de fausses détections	Taux de reconnaissance
RIE	86	14	86 %

**Tableau 4.6** Résultats de la méthode RIE.

#### **4.5. Conclusion:**

Pour la détection de visage, la première implémentation donne de bons résultats de détection et fonctionne avec des images couleur ou à niveaux de gris, son inconvénient majeur est le temps d'exécution requis pour détecter des visages de différentes tailles, la deuxième implémentation est beaucoup plus rapide que la première mais se limite aux images couleur.

Pour la reconnaissance de visage par la méthode des visages propres, on constate que:

- L'augmentation du nombre d'individus différents dans les ensembles de test et d'apprentissage permet d'améliorer le taux de reconnaissance.
- Plus il y a d'images par individu dans l'ensemble d'apprentissage et plus le taux de reconnaissance augmente.
- Un rapport élevé du nombre d'images dans l'ensemble d'apprentissage au nombre d'images dans l'ensemble de test permet d'avoir un meilleur taux de reconnaissance.
- La distance de Manhattan est celle qui donne le plus grand taux de reconnaissance parmi les autres distances.

On note toutefois que plus on donne de grandes valeurs à ces paramètres (nombre d'individus, nombre d'images par individu et rapport du nombre d'images d'apprentissage au nombre d'images de test) plus le taux de reconnaissance s'améliore, mais au détriment des performances du système qui devient plus lent et qui fait face à des données de plus en plus volumineuses (matrice de covariance, vecteurs propres,...).

Pour la méthode des visages propres des images réduites, on constate qu'elle apporte deux améliorations essentielles par rapport à la technique classique des visages propres. D'une part, elle donne un meilleur taux de reconnaissance, d'autre part, elle permet de résoudre (relativement) le problème de la grande dimension de la matrice de covariance des visages propres.

# **Annexe A:**

# **Mesures de dissimilitude**

### **A.1 Mesures de dissimilitude:**

Les formes ou les objets (images dans notre cas) analysés sont généralement représentés par des vecteurs (ou matrices) de mesures (couleur ou intensité de chaque pixel). Plusieurs techniques étudiées requièrent une certaine mesure de dissimilitude ou de distance entre deux vecteurs représentant des objets ou des formes, quoique parfois, ces données puissent se présenter directement sous forme de matrices de dissimilitude.

Une classe particulière de fonctions de dissimilitude appelée *coefficients de dissimilitude* doit satisfaire les conditions suivantes. Si  $d_{rs}$  est la dissimilitude entre deux objets  $r$  et  $s$ , alors:

$$\begin{aligned} d_{rs} &\geq 0 && \forall r, s \\ d_{rr} &= 0 && \forall r \\ d_{rs} &= d_{sr} && \forall r, s \end{aligned}$$

La condition de symétrie n'est pas toujours satisfaite par quelques fonctions de dissimilitude. Si la dissimilitude entre deux places dans une ville est la distance parcourue en route entre elles, alors cette distance peut être grande dans un sens que dans l'autre à cause des sens interdits dans cette ville.

Les mesures de dissimilitude peuvent être transformées en mesures de similitude via diverses transformations, par exemple:

$$s_{ij} = \frac{1}{1+d_{ij}} \quad \text{ou} \quad s_{ij} = c - d_{ij}$$

Où  $c$  est une constante et  $s_{ij}$  est la similitude entre les objets  $i$  et  $j$ .

Si en plus des trois conditions citées ci-dessus, la mesure de dissimilitude satisfait l'inégalité triangulaire:

$$d_{rr} + d_{ts} \geq d_{rs} \quad \forall r, s, t$$

Alors, la mesure de dissimilitude est dite *métrique* est le terme *distance* est utilisé.

#### **A.1.1. Variables numériques:**

Plusieurs mesures de dissimilitude ont été proposées pour les variables numériques. Le tableau A.1 donne quelques unes des plus utilisées. Le choix d'une métrique particulière dépend de l'application. Si on laisse à part les considérations concernant le temps de calcul, la métrique qui donnera la meilleure performance (peut être en terme d'erreur de classification) sera adoptée.

**Tableau A.1** Mesures de dissimilitude pour variables numériques (entre  $x$  et  $y$ )

Mesure de dissimilitude	Forme mathématique
Distance Euclidienne	$d_e = \left\{ \sum_{i=1}^p (x_i - y_i)^2 \right\}^{\frac{1}{2}}$
Distance de Manhattan (City-block)	$d_{cb} = \sum_{i=1}^p  x_i - y_i $
Distance de Chebyshev	$d_{ch} = \max_i  x_i - y_i $
Distance de Minkowski d'ordre $m$	$d_M = \left\{ \sum_{i=1}^p  x_i - y_i ^m \right\}^{\frac{1}{m}}$
Distance quadratique	$d_q = \sum_{i=1}^p \sum_{j=1}^p (x_i - y_i) Q_{ij} (x_i - y_i)$ <i>Q définie positive</i>
Distance de Canberra	$d_{ca} = \sum_{i=1}^p \frac{ x_i - y_i }{x_i + y_i}$
Distance non linéaire	$d_n = \begin{cases} H & \text{si } d_e > D \\ 0 & \text{si } d_e \leq D \end{cases}$
Séparation angulaire	$d_{as} = \frac{\sum_{i=1}^p x_i y_i}{\left[ \sum_{i=1}^p x_i^2 \sum_{i=1}^p x_j^2 \right]^{\frac{1}{2}}}$

#### **A.1.1.1. Distance Euclidienne:**

$$d_e = \left\{ \sum_{i=1}^p (x_i - y_i)^2 \right\}^{\frac{1}{2}}$$

Les contours de distance Euclidienne égale par rapport à un point arbitraire forment des hyper sphères (cercles en deux dimensions). Cette distance a la propriété (parfois indésirable) de donner une grande emphase aux différences larges sur une seule variable.

Quoiqu'on a tendance à utiliser des mesures de dissimilitude qui sont des métriques, certaines applications n'exigent pas la condition de métrique ci-dessus. Donc, dans certains cas, une fonction monotone de la distance Euclidienne, qui appartient toujours à la classe des coefficients de dissimilitude, mais qui n'est pas nécessairement une métrique sera suffisante. Par exemple, le carré de la distance Euclidienne est un coefficient de dissimilitude mais pas une métrique.

#### **A.1.1.2. Distance de Manhattan (City-block):**

$$d_{cb} = \sum_{i=1}^p |x_i - y_i|$$

Connue aussi sous le nom de *distance de valeur absolue* (*absolute value distance*), cette métrique utilise une méthode de calcul qui est convenable pour trouver les distances entre des points dans une ville constituée d'intersections de rues (d'où le nom utilisé). Les contours de distance de Manhattan égale par rapport à un point arbitraire forment des losanges ou des diamants en deux dimensions.

La distance de Manhattan est un peu plus rapide et simple à calculer par rapport à la distance Euclidienne, ce qui favorise son utilisation dans des applications où la vitesse d'exécution est critique.

### **A.1.1.3. Distance de Chebyshev:**

$$d_{ch} = \max_i |x_i - y_i|$$

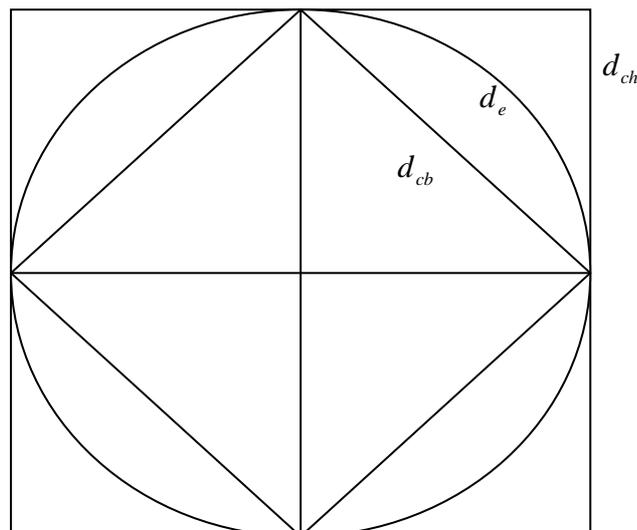
La distance de Chebyshev ou la *distance de valeur maximale* est utilisée dans des cas où la vitesse d'exécution est tellement critique que le temps impliqué dans le calcul de la distance Euclidienne est inacceptable. La distance de Chebyshev, comme celle de Manhattan, examine la valeur absolue des différences entre les éléments d'une paire de vecteurs. Les contours de distance de Chebyshev égale par rapport à un point arbitraire forment des losanges ou des carreaux en deux dimensions.

Si on a besoin d'une approximation de la distance Euclidienne mais avec moins de calculs, une solution serait d'utiliser soit la distance de Manhattan, soit la distance de Chebyshev, une meilleure solution consisterait à utiliser une combinaison de ces deux distances.

$$d = \max\left(\frac{2}{3}d_{cb}, d_{ch}\right)$$

En deux dimensions, les contours de distances égales forment des octogones.

La figure A.1 trace les contours de distances égales dans  $\square^2$  pour la distance Euclidienne, la distance de Manhattan et la distance de Chebyshev.



**Figure A.1** Contours de distance égale

**A.1.1.4. Distance de Minkowski:**

C'est une forme plus générale de la distance Euclidienne et de la distance de Manhattan. La distance de Minkowski d'ordre  $m$  est:

$$d_M = \left\{ \sum_{i=1}^p |x_i - y_i|^m \right\}^{\frac{1}{m}}$$

La distance de Minkowski d'ordre 1 est la distance de Manhattan, d'ordre 2 est la distance Euclidienne. Les contours de distance de Minkowski égale par rapport à un point arbitraire forment des cercles carrelés qui s'arrondissent au fur et à mesure que  $m$  grandit. Le choix de la valeur de  $m$  dépend du degré d'emphase que l'on souhaite attribuer aux différences larges: les grandes valeurs de  $m$  donneront progressivement plus d'emphase aux larges différences  $|x_i - y_i|$ , lorsque  $m$  tend vers l'infini la distance tend vers la distance de Chebyshev (des contours sous forme de carreaux).

**A.1.1.5. Distance quadratique:**

$$d_q = \sum_{i=1}^p \sum_{j=1}^p (x_i - y_i) Q_{ij} (x_i - y_i)$$

Un choix de  $Q$  est la matrice de covariance intra classe. Cette distance est souvent référée à la *distance de Mahalanobis*.

**A.1.1.6. Distance de Canberra:**

$$d_{ca} = \sum_{i=1}^p \frac{|x_i - y_i|}{x_i + y_i}$$

La distance de Canberra est la somme d'une série de fractions et elle est convenable pour des variables prenant toujours des valeurs non négatives. Si  $x_i$  et  $y_i$  sont tous les deux égaux à zéro, cette distance est considérée comme égale à zéro. Si un seul des deux termes est égal à zéro, cette distance est égale à un.

**A.1.1.7. Distance non linéaire:**

$$d_n = \begin{cases} H & \text{si } d_e > D \\ 0 & \text{si } d_e \leq D \end{cases}$$

Où  $D$  est un seuil et  $H$  une constante.

**A.1.1.8. Séparation angulaire:**

$$d_{as} = \frac{\sum_{i=1}^p x_i y_i}{\left[ \sum_{i=1}^p x_i^2 \sum_{i=1}^p x_j^2 \right]^{\frac{1}{2}}}$$

La séparation angulaire est une similitude plutôt qu'une dissimilitude qui mesure l'angle entre deux vecteurs unitaires dans la direction des deux vecteurs qui représentent les deux objets (formes) d'intérêt. Cette distance est appropriée lorsque seules les amplitudes relatives sont importantes.

Le choix d'une mesure particulière dépend de l'application en question et de plusieurs facteurs (distribution de données, temps et vitesse de calculs,...). Il est impossible de faire des recommandations et les études effectuées ont été empiriques, mais la mesure que l'on doit adopter est celle dont on croit capturer les différences essentielles entre les objets ou les formes.

### **A.1.2. Variables nominales et ordinales:**

Les variables nominales et ordinales sont généralement représentées par un ensemble de variables binaires. Par exemple, une variable nominale avec  $s$  états est représentée avec  $s$  variables binaires. Si elle est à l'état  $m$ , donc chacune des  $s$  variables binaires a une valeur de zéro sauf la  $m$  ème variable, qui a une valeur de un. La dissimilitude entre deux objets peut être obtenue en sommant les contributions des variables individuelles.

En ce qui concerne les variables ordinales, la contribution d'une variable individuelle à la dissimilitude entre deux objets ne dépend pas simplement du fait que cette variable prend la même valeur pour ces deux objets. Si la contribution d'une variable à l'état  $m$  et d'une autre à l'état  $l$  ( $m < l$ ) est  $\delta_{ml}$ , donc on doit avoir:

$$\begin{aligned} \delta_{ml} &\geq \delta_{ms} && \text{pour } s < l \\ \delta_{ml} &\geq \delta_{sl} && \text{pour } s > m \end{aligned}$$

Ceci est,  $\delta_{ml}$  est monotone tout au long des lignes (de haut en bas) et des colonnes (de gauche à droite) de la demi matrice (obtenue en négligeant les éléments sous la diagonale) de distance entre les états ( $\delta_{14} > \delta_{13} > \delta_{12}$  etc.;  $\delta_{14} > \delta_{24} > \delta_{34}$ ). Les valeurs choisies pour  $\delta_{ml}$  dépendent fortement du problème en question. Par exemple, on peut avoir des variables décrivant des fruits avec les valeurs: *court, long ou très long*. On veut que la dissimilitude entre un arbre dont les fruits sont très longs et un autre avec des fruits courts soit plus grande que la dissimilitude entre un arbre dont les fruits sont longs et un autre avec des fruits courts (tous les autres attribues sont de valeurs égales). Un codage numérique de 1,2 et 3 conviendra dans notre cas, comme le fera aussi le codage 1,10 et 100.

### **A.1.3. Variables binaires:**

Plusieurs mesures de dissimilitude ont été proposées pour les variables binaires. Pour les vecteurs de variables binaires  $x$  et  $y$ , ces mesures peuvent être exprimées en termes des quantités  $a, b, c$  et  $d$  où:

- $a$  est égale au nombre d'occurrences de  $x_i = 1$  et  $y_i = 1$
- $b$  est égale au nombre d'occurrences de  $x_i = 0$  et  $y_i = 1$
- $c$  est égale au nombre d'occurrences de  $x_i = 1$  et  $y_i = 0$
- $d$  est égale au nombre d'occurrences de  $x_i = 0$  et  $y_i = 0$

Le tableau A.2 récapitule ce qu'on vient de dire, notons que  $a+b+c+d = p$ , le nombre total des variables (attribuées). Il est plus pratique de définir une mesure de similitude plutôt qu'une mesure de dissimilitude. Le tableau A.3 donne quelques unes des mesures de similitude les plus utilisées.

**Tableau A.2** Cooccurrence pour les variables binaires

		$x_i$	
		1	0
$y_i$	1	$a$	$b$
	0	$c$	$d$

**Tableau A.3** Mesures de similitude pour les données binaires

Mesure de similitude	Forme mathématique
Simple matching coefficient	$d_{sm} = \frac{a+d}{a+b+c+d}$
Russell et Rao	$d_{rr} = \frac{a}{a+b+c+d}$
Jaccard	$d_j = \frac{a}{a+b+c}$
Czekanowski	$d_{cz} = \frac{2a}{2a+b+c}$

#### **A.1.3.1. Simple matching coefficient:**

Cette mesure représente la proportion de variables pour laquelle deux variables ont la même valeur. L'insatisfaction avec cette mesure est dans le terme  $d$  qui représente deux absences conjointes. Le fait que deux sites dans une étude écologique manquent tous les deux de quelque chose n'augmente pas leur similitude. La mesure de dissimilitude définie par  $d_{xy} = 1 - s_{xy} = (b+c)/p$  est proportionnelle au carré de la distance Euclidienne,  $(b+c)$ , qui n'est que la *distance de Hamming* et théorie de communication.

#### **A.1.3.2. Russell et Rao:**

Cette mesure n'implique pas le terme  $d$  dans le numérateur et est appropriée dans certaines circonstances. La quantité  $1 - s_{xy}$  ne peut pas être considérée comme un coefficient de dissimilitude puisque la dissimilitude entre un objet et lui-même n'est pas nécessairement égale à zéro.

**A.1.3.3. Jaccard:**

Cette mesure n'implique pas le terme  $d$  carrément et est utilisée extensivement par les écologistes. Le terme  $d_{xy} = 1 - s_{xy}$  est un coefficient de dissimilitude métrique.

**A.1.3.4. Czekanowski:**

Cette mesure est similaire à la mesure de Jaccard excepté que les coïncidences ont un poids double.

Plusieurs autres coefficients qui gèrent les absences conjointes ont été proposés dans la littérature.

**A.2. Sommaire:**

On a listé quelques unes des mesures de similitude pouvant être trouvées dans la littérature de la reconnaissance des formes et de la théorie de la classification. Un coefficient de similitude plus général entre deux objets  $x$  et  $y$  contenant des variables de types divers a été proposé par [Gower (1971)]. Bien sûr, il n'y a pas de chose appelée meilleure mesure. Quelques unes peuvent être plus appropriées pour certaines tâches que d'autres. Donc, on ne peut pas faire de recommandations. Cependant, on peut considérer les points suivants lors du choix d'une mesure: (1) la simplicité et la facilité de compréhension; (2) la facilité d'implémentation; (3) les exigences en terme de vitesse d'exécution; (4) la connaissance concernant les données disponibles.

Pour plus de détails sur les mesures de dissimilitude, le lecteur est dirigé vers les sources suivantes: [Gower 1971 ; Gordon 1999 ; Diday et Simon 1976 ; Cormack 1971 ; Clifford et Stephenson 1975 ; Kittler 1975 ; Kittler 1986 ; Chen 1976 ; Devijver et Kittler 1982].

# **Annexe B:**

# **Rappels d'Algèbre linéaire**

**B.1. Définitions et propriétés fondamentales:**

A travers cette thèse, on assume que le lecteur est familier avec les opérations standard sur les vecteurs et les matrices. Cette annexe est incluse comme une référence pour la terminologie et la notation utilisées dans cette thèse. Toutes les propriétés des matrices seront évoquées sans démonstration. Les démonstrations peuvent être trouvées dans n'importe quel livre d'algèbre linéaire (voir références plus loin). On donnera aussi des informations sur des sources de logiciels pour effectuer quelques unes des opérations standard sur les matrices.

Etant donnée une matrice  $m \times n$ , on dénote l'élément de la  $i^{\text{ème}}$  ligne et de la  $j^{\text{ème}}$  colonne par  $a_{ij}$ . La *transposée* d'une matrice  $A$  est notée  $A^T$ , et on note que:

$$(AB)^T = B^T A^T$$

La matrice carrée  $A$  est *symétrique* si  $a_{ij} = a_{ji} \forall i, j$ .

La *trace* d'une matrice carrée  $A$ , notée  $Tr\{A\}$ , est la somme de ses éléments diagonaux,

$$Tr\{A\} = \sum_{i=1}^n a_{ii}$$

Elle satisfait  $Tr\{AB\} = Tr\{BA\}$  étant donné que  $AB$  est une matrice carrée, cependant  $A$  et  $B$  peuvent ne pas être des matrices carrées.

Le *déterminant* d'une matrice  $A$ , noté  $|A|$ , est la somme

$$|A| = \sum_{j=1}^n a_{ij} A_{ij} \quad \text{pour } i = 1, \dots, n$$

Où le *cofacteur*  $A_{ij}$  est le déterminant de la matrice formée par la suppression de la  $i^{\text{ème}}$  ligne et de la  $j^{\text{ème}}$  colonne de  $A$ , multiplié par  $(-1)^{i+j}$ . La matrice des cofacteurs  $C (c_{ij} = A_{ij})$  est appelée la *matrice adjointe* de  $A$ . Si  $A$  et  $B$  sont des matrices carrées du même ordre, alors  $|AB| = |A||B|$ .

L'*inverse* d'une matrice  $A$  est la matrice unique  $A^{-1}$  qui satisfait  $A^{-1}A = AA^{-1} = I$

Où  $I$  est la *matrice identité*. Si l'inverse existe, la matrice est dite *régulière* ou *non singulière*. Si l'inverse n'existe pas, elle est dite *singulière* et  $|A| = 0$ . On souhaite fréquemment utiliser les propriétés  $(A^T)^{-1} = (A^{-1})^T$ ,  $(AB)^{-1} = B^{-1}A^{-1}$  et si  $A$  est symétrique, donc  $A^{-1}$  l'est aussi.

Un ensemble de  $k$  vecteurs (de dimension égale) sont dits *linéairement dépendants* s'il existe un ensemble de scalaires  $c_1, \dots, c_k$  qui ne sont pas tous égaux à zéro tels que:

$$c_1 x_1 + \dots + c_k x_k = 0$$

S'il est impossible de trouver un tel ensemble  $c_1, \dots, c_k$ , alors les vecteurs  $x_1, \dots, x_k$  sont dits *linéairement indépendants*. Le *rang* d'une matrice est le nombre maximal de lignes (ou de colonnes) linéairement indépendants.

Une matrice  $n \times n$  est dite de *plein rang* si son rang est égal à  $n$ . Dans ce cas, son déterminant n'est pas nul, c'est-à-dire que son inverse existe. Pour une matrice rectangulaire d'ordre  $m \times n$ ,  $\text{rang}(A) \leq \min(m, n)$  et

$$\text{rang}(A) = \text{rang}(A^T) = \text{rang}(A^T A) = \text{rang}(A A^T)$$

Le rang d'une matrice reste inchangé après la pré ou la post multiplication par une matrice non singulière.

Une matrice carrée est dite *orthogonale* si  $A^T A = A A^T = I$ , ces lignes(colonnes) sont dits *ortho normaux* ( $x^T y = 0$  et  $x^T x = 1, y^T y = 1$  pour deux lignes ou colonnes différentes  $x$  et  $y$ ). Une matrice orthogonale représente une transformation linéaire qui préserve les distances et les angles et qui consiste en une rotation et/ou réflexion. Il est clair d'après la définition précédente qu'une matrice orthogonale est non singulière et que l'inverse d'une matrice orthogonale est égal à sa transposée  $A^{-1} = A^T$ . Aussi, il faut noter que le déterminant d'une matrice orthogonale est égal à  $\pm 1$  ( $-1$  indique une réflexion,  $+1$  indique une rotation pure).

Une matrice carrée est dite *définie positive* si la forme quadratique  $x^T A x > 0$  pour tout  $x \neq 0$ . Elle est dite *semi définie positive* si  $x^T A x \geq 0$  pour tout  $x \neq 0$ . Les matrices définies positives sont de plein rang.

Les *valeurs propres (ou racines caractéristiques)* d'une matrice  $A$  d'ordre  $p \times p$  sont les solutions de l'équation caractéristique

$$|A - \lambda I| = 0$$

Qui est un polynôme de degré  $p$ . Ainsi, il existe  $p$  solutions, qui sont notées  $\lambda_1, \dots, \lambda_p$ . Ces valeurs propres ne sont pas nécessairement distinctes et elles peuvent être réelles ou complexes. On associe à chaque valeur propre  $\lambda_i$  un vecteur propre  $u_i$  ayant la propriété

$$A u_i = \lambda_i u_i$$

Ces vecteurs propres ne sont pas uniques, puisque chaque multiple de  $u_i$  satisfait l'équation précédente. Donc, les vecteurs propres sont souvent normalisés de façon que  $u_i^T u_i = 1$ .

On donne ici quelques propriétés des valeurs et des vecteurs propres

1. Le produit des valeurs propres d'une matrice est égal à son déterminant, i.e.

$\prod_{i=1}^p \lambda_i = |A|$ . Ainsi, il s'ensuit que si aucune des valeurs propres n'est égale à zéro, alors l'inverse de la matrice  $A$  existe.

2. La somme des valeurs propres d'une matrice est égale à sa trace,

$$\sum_{i=1}^p \lambda_i = \text{Tr}\{A\}.$$

3. Si  $A$  est une matrice réelle et symétrique, alors ces valeurs et vecteurs propres sont réels aussi.

4. Si  $A$  est définie positive, ces valeurs propres sont toutes supérieures à zéro.

5. Si  $A$  est semi définie positive d'ordre  $m$ , alors il existe  $m$  valeurs propres non nulles et  $p - m$  valeurs propres nulles.

6. Chaque matrice réelle et symétrique possède un ensemble de vecteurs caractéristiques orthonormaux. Ainsi, la matrice  $U$  dont les colonnes sont les vecteurs propres d'une matrice réelle et symétrique ( $U = [u_1, \dots, u_p]$ ), est orthogonale,  $U^T U = U U^T = I$  et  $U^T A U = \Lambda$  où  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ , la matrice diagonale ayant comme éléments diagonaux les valeurs propres  $\lambda_i$ . Alternativement, on peut écrire

$$A = U \Lambda U^T = \sum_{i=1}^p \lambda_i u_i u_i^T$$

Si  $A$  est définie positive, alors  $A^{-1} = U \Lambda^{-1} U^T$ , où  $\Lambda^{-1} = \text{diag}(1/\lambda_1, \dots, 1/\lambda_p)$ .

L'équation générale des vecteurs propres symétriques (vecteurs propres généralisés) est

$$A u = \lambda B u$$

Où  $A$  et  $B$  sont des matrices réelles et symétriques, apparaît dans l'application de la LDA (*linear discriminant analysis*) et dans d'autres applications de la reconnaissance des formes. Si  $B$  est définie positive, alors l'équation précédente aura  $p$  vecteurs propres  $(u_1, \dots, u_p)$  qui sont orthonormaux par rapport à  $B$ , c'est-à-dire

$$u_i^T B u_j = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j \end{cases}$$

Et par conséquent

$$u_i^T A u_j = \begin{cases} 0 & \text{si } i \neq j \\ \lambda_j & \text{si } i = j \end{cases}$$

Ce qui peut s'écrire comme

$$U^T B U = I \quad U^T A U = \Lambda$$

Où  $U = [u_1, \dots, u_p]$  et  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ .

Finalement, dans cette section, on introduit quelques résultats à propos des dérivées. On note l'opérateur de dérivée partielle par

$$\frac{\partial}{\partial x} = \left( \frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_p} \right)^T$$

Ainsi, la dérivée de la fonction scalaire  $f$  du vecteur  $x$  est le vecteur

$$\frac{\partial f}{\partial x} = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_p} \right)^T$$

D'une façon similaire, la dérivée d'une fonction scalaire d'une matrice est notée  $\partial f / \partial A$ , où

$$\left[ \frac{\partial f}{\partial A} \right]_{ij} = \frac{\partial f}{\partial a_{ij}}$$

En particulier, on a

$$\frac{\partial |A|}{\partial A} = (\text{adj}(A))^T = |A| (A^{-1})^T \text{ si } A^{-1} \text{ existe}$$

Et pour une matrice symétrique

$$\frac{\partial}{\partial x} x^T A x = 2Ax$$

Enfin, une dérivée importante qui implique les traces de matrices est

$$\frac{\partial}{\partial A} (\text{Tr}\{A^T M A\}) = MA + M^T A$$

## **B.2 Sommaire:**

Dans cette annexe, on a introduit la terminologie nécessaire des matrices pour aborder cette thèse. On a été bref avec les détails et les démonstrations. Les démonstrations et autres détails peuvent être trouvés dans la plupart des livres d'algèbre linéaire, par exemple le livre de [Stewart 1973] fournit une très bonne introduction au calcul matriciel.

Pour une étude approfondie des notions de l'algèbre linéaire, le lecteur est dirigé vers les sources suivantes: [Stewart 1973 ; Press et autres 1992 ; Thisted 1988].

# **Conclusion et perspectives**

Dans ce mémoire, on a étudié les problèmes de la détection de visage et de la reconnaissance de visage par machine.

On a proposé deux méthodes pour la détection de visage à partir d'images statiques. La première méthode gère les images à niveaux de gris comme celles en couleur, de vues frontale ou presque frontales, elle procède tout d'abord par l'extraction des yeux qui sont les parties les plus sombres du visage humain, elle suppose ensuite la présence de candidats de visage pour chaque paire de segments analogues aux yeux ayant vérifié certaines contraintes anthropologiques relatives aux yeux humains, la vérification des candidats possibles de visages se fait par une classification dans l'espace propre de la PCA de ses formes en visage ou non visage selon leurs projections dans l'espace propre. Dans la deuxième méthode qui est plus rapide que la première mais qui ne gère que les images couleur, on a développé un modèle de segmentation dans l'espace des couleurs normalisées rgb, on ajouté des contraintes géométriques pour améliorer le résultat de la segmentation et on a vérifié la présence de visages par la recherche des segments analogues aux yeux.

Pour la reconnaissance de visage, on a implémenté la technique des visages propres (principal component analysis, PCA) qui permet de projeter les images des visages de grande dimension dans un espace de dimension réduite afin de rendre les calculs possibles sur un ordinateur, on a comparé aussi les performances des différents types de mesures de dissimilitude utilisés avec la PCA. La deuxième technique implémentée est celle des visages propres des images réduites (RIE) qui résout le problème de la grande dimension rencontré en PCA et qui donne de meilleurs résultats.

Dans le futur, nous espérons améliorer nos travaux en se penchant sur les problèmes suivants:

- La détection en temps réel du visage humain.
- La détection de visages humains de différentes vues (frontale, de profil,...)
- L'automatisation de la détection/reconnaissance de visage, c'est-à-dire le développement d'un système entièrement automatique qui regroupe à la fois un module de détection et un module de reconnaissance.

# Références

- [**Abdel-Mottaleb et Elgammal 1999**] M. Abdel-Mottaleb and A. Elgammal, 1999. Face detection in complex environments, in Proceedings International Conference on Image Processing.
- [**Albiol 2000**] A. Albiol, 2000. A simple and efficient face detection algorithm for video database applications, in Proceedings of the 2000 International Conference on Image Processing, p. TA07.09.
- [**Albiol et autres 1999**] A. Albiol, C. A. Bouman, and E. J. Delp, 1999. Face detection for pseudosemantic labeling in video databases, in Proceedings International Conference on Image Processing.
- [**Bartlett et autres 1998**] M. S. Bartlett, H. M. Lades, and T. Sejnowski, 1998. Independent component representation for face recognition. In Proceedings, SPIE Symposium on Electronic Imaging: Science and Technology. 528–539.
- [**Belhumeur et autres 1997**] P. N. Belhumeur and D. J. Kriegman, 1997. What is the set of images of an object under all possible lighting conditions? In Proceedings, IEEE Conference on Computer Vision and Pattern Recognition. 52–58.
- [**Bell et Sejnowski 1995**] A. J. Bell and T. J. Sejnowski, 1995. An information maximisation approach to blind separation and blind deconvolution. *Neural Computation* 7, 1129–1159.
- [**Bell et Sejnowski 1997**] A. J. Bell and T. J. Sejnowski, 1997. The independent components of natural scenes are edge filters. *Vis. Res.* 37, 3327–3338.
- [**Ben-Yacoub et autres 1999**] S. Ben-Yacoub, B. Fasel, and J. Lüttin, 1999. Fast face detection using MLP and FFT, in Proceedings Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA).
- [**Biederman et Kalocsai 1998**] I. Biederman and P. Kalocsai, 1998. Neural and psychophysical analysis of object and face recognition. In *Face Recognition: From Theory to Applications*, H. Wechsler, P. J. Phillips, V. Bruce, F. F. Soulie, and T. S. Huang, Eds. Springer-Verlag, Berlin, Germany, 3–25.
- [**Biométrie**] [www.biometrie-online.net](http://www.biometrie-online.net)
- [**Blanz et Vetter 1999**] V. Blanz and T. Vetter, 1999. A Morphable model for the synthesis of 3D faces. In Proceedings, SIGGRAPH'99, 187–194.
- [**Bledsoe 1964**] W. W. Bledsoe, 1964. The model method in facial recognition. Tech. rep. PRI:15, Panoramic research Inc., Palo Alto, CA.
- [**Bronstein et autres 2003**] A. Bronstein, M. Bronstein, E. Gordon, and R. Kimmel, 2003. 3D face recognition using geometric invariants. In Proceedings, International Conference on Audio- and Video-Based Person Authentication.
- [**Bruce 1988**] V. Bruce, 1988. *Recognizing faces*, Lawrence Erlbaum Associates, London, U.K.
- [**Brunelli et Poggio 1993**] R. Brunelli and T. Poggio, 1993. Face recognition: Feature versus templates, *IEEE Trans. Pattern Anal. Mach. Intell.* 15, 1042–1052.
- [**Bruner and Tagiuri 1954**] I. S. Bruner and R. Tagiuri, 1954. The perception of people. In *Handbook of Social Psychology*, Vol. 2, G. Lindzey, Ed., Addison-Wesley, Reading, MA, 634–654.
- [**Buhmann et autres 1990**] J. Buhmann, M. Lades, and C. V. D. Malsburg, 1990. Size and distortion invariant object recognition by hierarchical graph matching. In Proceedings, International Joint Conference on Neural Networks. 411–416.
- [**Burel et Carel 1994**] G. Burel and D. Carel, 1994. Detection and localization of faces on digital images, *Pattern Recog. Lett.* 15, 963–967.
- [**Cai et Goshtasby 1999**] J. Cai and A. Goshtasby, 1999. Detecting human faces in color images, *Image Vision Comput.* 18, 63–75.

- [**Chellappa et autres 1995**] R. Chellappa, C. L. Wilson, and S. Sirohey, 1995. Human and machine recognition of faces: A survey, *Proc. IEEE* 83, 5.
- [**Chen 1976**] C.H. Chen, 1976. On information and distance measures, error bounds, and feature selection. *Information Sciences*, 10:159–173.
- [**Chen et Chiang 1997**] C. Chen and S. P. Chiang, 1997. Detection of human faces in color images, *IEE Proc. Vision Image Signal Process.* 144, 384–388.
- [**Chichizola et autres 2005**] F. Chichizola, L. De Giusti, and M. Naiouf, 2005. Face Recognition: Reduced Image Eigenfaces Method. 47<sup>th</sup> International Symposium ELMAR-2005, 08-10 June 2005, Zadar, Croatia, 159-161.
- [**Choi et autres 1999**] J. Choi, S. Kim, and P. Rhee, 1999. Facial components segmentation for extracting facial feature, in *Proceedings Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*.
- [**Clifford et Stephenson 1975**] H.T. Clifford and W. Stephenson, 1975. *An Introduction to Numerical Classification*. Academic Press, New York.
- [**Colmenarez et autres 1999**] A. J. Colmenarez, B. Frey, and T. S. Huang, 1999. Detection and tracking of faces and facial features, in *Proceedings International Conference on Image Processing*.
- [**Colmenarez et Huang 1996**] A. J. Colmenarez and T. S. Huang, 1996. Maximum likelihood face detection, in *IEEE Proc. of 2nd Int. Conf. on Automatic Face and Gesture Recognition*, Vermont, 1996, pp. 222–224.
- [**Colmenarez et Huang 1997**] A. J. Colmenarez and T. S. Huang, 1997. Face detection with information-based maximum discrimination, in *IEEE Proc. of Int. Conf. on Computer Vision and Pattern Recognition*.
- [**Colmenarez et Huang 1998**] A. J. Colmenarez and T. S. Huang, 1998. Pattern detection with information-based maximum discrimination and error bootstrapping, in *Proc. of International Conference on Pattern Recognition*.
- [**Cootes et autres 1995**] T. Cootes, C. Taylor, D. Cooper, and J. Graham, 1995. Active shape models—their training and application. *Comput. Vis. Image Understand.* 61, 18–23.
- [**Cootes et autres 2000**] T. Cootes, K. Walker and C. Taylor, 2000. Viewbased active appearance models. In *Proceedings, International Conference on Automatic Face and Gesture Recognition*.
- [**Cootes et autres 2001**] T. F. Cootes, G. J. Edwards, and C. J. Taylor, 2001. Active appearance models. *IEEE Trans. Patt. Anal. Mach. Intell.* 23, 681–685.
- [**Cormack 1971**] R.M. Cormack, 1971. A review of classification (with discussion). *Journal of the Royal Statistical Society Series A*, 134:321–367.
- [**Cox et autres 1996**] I. J. Cox, J. Ghosn, and P. N. Yianilos, 1996. Feature-based face recognition using mixture distance. In *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*. 209–216.
- [**Craw et autres 1987**] I. Craw, H. Ellis, and J. R. Lishman, 1987. Automatic extraction of face-feature, *Pattern Recog. Lett.* Feb. 1987, 183–187.
- [**Craw et Cameron 1996**] I. Craw and P. Cameron, 1996. Face recognition by computer. In *Proceedings, British Machine Vision Conference*. 489–507.
- [**Crowley et Berard 1997**] J. L. Crowley and F. Berard, 1997. Multi-model tracking of faces for video communications, in *IEEE Proc. of Int. Conf. on Computer Vision and Pattern Recognition*, Puerto Rico, Jun. 1997.
- [**CVL FACE DATABASE**] CVL FACE DATABASE: <http://www.lrv.fri.uni-lj.si/facedb.html>
- [**Dai et Nakano 1996**] Y. Dai and Y. Nakano, 1996. Face-texture model based on sgld and its application, *Pattern Recog.* 29, 1007–1017.

- [**Darrell et autres 1998**] T. Darrell, G. Gordon, M. Harville, and J. Woodfill, 1998. Integrated person tracking using stereo, color, and pattern detection, in IEEE Proc. of Int. Conf. on Computer Vision and Pattern Recognition.
- [**Darwin 1972**] C. Darwin, 1972. *The Expression of the Emotions in Man and Animals*. John Murray, London, U.K.
- [**De Silva et autres 1995**] L. C. De Silva, K. Aizawa, and M. Hatori, 1995. Detection and tracking of facial features by using a facial feature model and deformable circular template, IEICE Trans. Inform. Systems E78–D(9), 1195–1207.
- [**Demirel et autres 1996**] H. Demirel, T. J. Clarke, and P. J. K. Cheung, 1996. Adaptive automatic facial feature segmentation, in IEEE Proc. of 2nd Int. Conf. on Automatic Face and Gesture Recognition, Vermont, Oct. 1996, pp. 277–282.
- [**Dempster et autres 1977**] A. P. Dempster, N. M. Laird, and D. B. Rubin, 1977. Maximum likelihood from incomplete data via the EM algorithm, *J. Roy. Statist. Soc.* 39, 1–39.
- [**Devijver et Kittler 1982**] P.A. Devijver and J. Kittler, 1982. *Pattern Recognition, A Statistical Approach*. Prentice Hall, London.
- [**Diday et Simon 1976**] E. Diday and J.C. Simon, 1976. Clustering analysis. In K.S. Fu, ed., *Digital Pattern Recognition*, pp. 47–94. Springer-Verlag, Berlin.
- [**Duchnowski et autres 1995**] P. Duchnowski, M. Hunke, D. Busching, U. Meier, and A. Waibel, 1995. Toward movement invariant automatic lip-reading and speech recognition, in Proc. Int. Conf. on Acoustic, Speech and Signal Processing.
- [**Duda et Hart 1973**] R. O. Duda and P. E. Hart, 1973. *Pattern Classification and Scene Analysis*, Wiley, New York.
- [**Duta et Jain 1998**] N. Duta and A. K. Jain, 1998. Learning the human face concept from black and white images, in Proc. of International Conference on Pattern Recognition.
- [**Ekman 1998**] P. Ed. Ekman, 1998. *Charles Darwin's The Expression of the Emotions in Man and Animals, Third Edition, with Introduction, Afterwords and Commentaries by Paul Ekman*. Harper-Collins/Oxford University Press, New York, NY/London, U.K.
- [**Etemad et Chellappa 1997**] K. Etemad and R. Chellappa, 1997. Discriminant analysis for recognition of human face images. *J. Opt. Soc. Am. A* 14, 1724–1733.
- [**Feraud et autres 1997**] R. Feraud, O. Bernier, and D. Collobert, 1997. A constrained generative model applied to face detection, *Neural Process. Lett.* 5, 73–81.
- [**Feraud et autres 2000**] R. Feraud, O. Bernier, J.-E. Viallet, and M. Collobert, 2000. A fast and accurate face detector for indexation of face images, in Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition.
- [**Fisher 1936**] R. A. Fisher, 1936. The use of multiple measurements in taxonomic problems, *Ann. Eugenics* 7, 179–188.
- [**Fisher 1938**] R. A. Fisher, 1938. The statistical utilization of multiple measurements. *Ann. Eugen.* 8, 376–386.
- [**Fouad et autres 2000**] M. Fouad, A. Darwish, S. Shaheen, and F. Bayoumi, 2000. Mode-based human face detection in unconstrained scenes, in Proceedings of the 2000 International Conference on Image Processing, p. TA07.06.
- [**Fukunaga 1989**] K. Fukunaga, 1989. *Statistical Pattern Recognition*, Academic Press, New York, NY.
- [**Galton 1888**] F. Galton, 1888. Personal identification and description. *Nature*, (June 21), 173–188.

- [**Garcia et Tziritas 1999**] C. Garcia and G. Tziritas, 1999. Face detection using quantized skin color regions, merging and wavelet packet analysis, *IEEE Trans. Multimedia* 1, 264–277.
- [**Ghahramani et Hinton 1996**] Z. Ghahramani and G. 1996. Hinton, The EM Algorithm for Mixtures of Factor Analyzers, Technical Report CRG-TR-96-1, Dept. of Computer Science, University of Toronto, 1996.
- [**Gordon 1991**] G. Gordon, 1991. Face recognition based on depth maps and surface curvature. In *SPIE Proceedings, Vol. 1570: Geometric Methods in Computer Vision*. SPIE Press, Bellingham, WA 234–247.
- [**Gordon 1999**] A.D. Gordon, 1999. *Classification*, 2nd edn., Chapman & Hall/CRC, Boca Raton, FL.
- [**Govindaraju 1996**] V. Govindaraju, 1996. Locating human faces in photographs, *Int. J. Comput. Vision* 19.
- [**Gower 1971**] J.C. Gower, 1971. A general coefficient of similarity and some of its properties. *Biometrics*, 27:857–874.
- [**Graf et autres 1996**] H. P. Graf, E. Cosatto, D. Gibson, E. Petajan, and M. Kocheisen, 1996. Multi-modal system for locating heads and faces, in *IEEE Proc. of 2nd Int. Conf. on Automatic Face and Gesture Recognition*, Vermont, Oct. 1996, pp. 277–282.
- [**Graf et autres 2000**] H. Graf, E. Cosatto, and T. Ezzat, 2000. Face analysis for the synthesis of photo-realistic talking heads, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition*.
- [**Gu et autres 2001**] L. Gu, S. Z. Li, and H. J. Zhang, 2001. Learning probabilistic distribution model for multiview face detection. In *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*.
- [**Gu et Li 2000**] Q. Gu and S. Z. Li, 2000. Combining feature optimization into neural network based face detection, in *Proceedings of the 15th International Conference on Pattern Recognition*, Vol. II, p. 4A.
- [**Hallinan 1991**] P.W. Hallinan, 1991. Recognizing human eyes. In *SPIE Proceedings, Vol. 1570: Geometric Methods In Computer Vision*. 214–226.
- [**Hallinan 1994**] P. W. Hallinan, 1994. A low-dimensional representation of human faces for arbitrary lighting conditions. In *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*. 995–999.
- [**Han et autres 2000**] C. C. Han, H. Y. M. Liao, G. J. Yu, and L. H. Chen, 2000. Fast face detection via morphology-based pre-processing, *Pattern Recog.* 33.
- [**Heisele et autres 2001**] B. Heisele, T. Serre, M. Pontil, and T. Poggio, 2001. Component-based face detection. In *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*.
- [**Herpers et autres 1995**] R. Herpers, H. Kattner, H. Rodax, and G. Sommer, 1995. Gaze: An attentive processing strategy to detect and analyze the prominent facial regions, in *IEEE Proc. of Int. Workshop on Automatic Face- and Gesture-Recognition*, Zurich, Switzerland, Jun. 1995, pp. 214–220.
- [**Herpers et autres 1996**] R. Herpers, M. Michaelis, K.-H. Lichtenauer, and G. Sommer, 1996. Edge and keypoint detection in facial regions, in *IEEE Proc. of 2nd Int. Conf. on Automatic Face and Gesture Recognition*, Vermont, Oct. 1996, pp. 212–217.
- [**Herpers et autres 1999**] R. Herpers, G. Verghese, K. Derpanis, R. McCready, J. MacLean, A. Jepson, and J. K. Tsotsos, 1999. Detection and tracking of faces in real environments, in *Proc. IEEE International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*.

- [**Hjelmas et Low 2001**] E. Hjelmas and B. K. Low, 2001. Face detection: A Survey. *Comput. Vis. Image Understand.* 83, 236–274.
- [**Holst 2000**] G. Holst, 2000. Face detection by facets: Combined bottom-up and top-down search using compound templates, in *Proceedings of the 2000 International Conference on Image Processing*, p. TA07.08.
- [**Hongo et autres 2000**] H. Hongo, M. Ohya, M. Yasumoto, Y. Niwa, and K. Yamamoto, 2000. Focus of attention for face and hand gesture recognition using multiple cameras, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition*.
- [**Hoogenboom et Lew 1996**] R. Hoogenboom and M. Lew, 1996. Face detection using local maxima, in *IEEE Proc. of 2nd Int. Conf. on Automatic Face and Gesture Recognition*, Vermont, Oct. 1996, pp. 334–339.
- [**Hotta et autres 1998**] K. Hotta, T. Kurita, and T. Mishima, 1998. Scale invariant face detection method using higher-order local autocorrelation features extracted from log-polar image, in *IEEE Proc. of 3rd Int. Conf. on Automatic Face and Gesture Recognition*.
- [**Hu 1962**] M. K. Hu, 1962. Visual pattern recognition by moment invariants, *IRE Transactions on Information Theory* 8, 179–187.
- [**Huang et autres 1996**] J. Huang, S. Gutta, and H. Wechsler, 1996. Detection of human faces using decision trees, in *IEEE Proc. of 2<sup>nd</sup> Int. Conf. on Automatic Face and Gesture Recognition*, Vermont, 1996.
- [**Huang et autres 2003**] J. Huang, , B. Heisele, and V. Blanz, 2003. Component-based face recognition with 3D morphable models. In *Proceedings, International Conference on Audio- and Video-Based Person Authentication*.
- [**Hunke et Waibel 1994**] M. Hunke and A. Waibel, 1994. Face locating and tracking for human-computer interaction, in *28th Asilomar Conference on Signals, Systems and Computers*, Monterey, CA.
- [**Jacquin et Eleftheriadis 1995**] A. Jacquin and A. Eleftheriadis, 1995. Automatic location tracking of faces and facial features in video sequences, in *IEEE Proc. of Int. Workshop on Automatic Face- and Gesture-Recognition*, Zurich, Switzerland, Jun. 1995.
- [**Jebara et Pentland 1997**] T. S. Jebara and A. Pentland, 1997. Parametrized structure from motion for 3D adaptive feedback tracking of faces, in *IEEE Proc. of Int. Conf. on Computer Vision and Pattern Recognition*, Puerto Rico.
- [**Jesorsky et autres 2001**] O. Jesorsky, K. Kirchberg, and R. Frischholz, 2001. Robust face detection using the Hausdorff distance, in: J. Bigun, F. Smeraldi (Eds.), *Audio and Video based Person Authentication—AVBPA 2001*, Springer, Berlin, pp. 90–95.
- [**Jiang et autres 2000**] X. Jiang, M. Binkert, B. Achermann, and H. Bunke, 2000. Towards detection of glasses in facial images, *Pattern Anal. Appl.* 3, 9–18.
- [**Jing et Mariani 2000**] Z. Jing and R. Mariani, 2000. Glasses detection and extraction by deformable contour, in *Proceedings of the 15<sup>th</sup> International Conference on Pattern Recognition*, Vol. II, p. 4B.
- [**Jordão et autres 1999**] L. Jordão, M. Perrone, and J. P. Costeira, 1999. Active face and feature tracking, in *Proceedings of the 10<sup>th</sup> International Conference on Image Analysis and Processing*.
- [**Juell et Marsh 1996**] P. Juell and R. Marsh, 1996. A hierarchical neural network for human face detection, *Pattern Recog.* 29, 781–787.
- [**Kanade 1973**] T. Kanade, 1973. *Computer recognition of human faces*. Birkhauser, Basel, Switzerland, and Stuttgart, Germany.

- [**Kapfer et Benois-Pineau 1997**] M. Kapfer and J. Benois-Pineau, 1997. Detection of human faces in color image sequences with arbitrary motions for very low bit-rate videophone coding, *Pattern Recog. Lett.* 18.
- [**Karlekar et Desai 1999**] J. Karlekar and U. B. Desai, 1999. Finding faces in color images using wavelet transform, in *Proceedings of the 10th International Conference on Image Analysis and Processing*.
- [**Kawaguchi et autres 2000**] T. Kawaguchi, D. Hidaka, and M. Rizon, 2000. Robust extraction of eyes from face, in *Proceedings of the 15<sup>th</sup> International Conference on Pattern Recognition*, Vol. I, p. 3A.
- [**Kawato et Ohya 2000**] S. Kawato and J. Ohya, 2000. Real-time detection of nodding and head-shaking by directly detecting and tracking the “between-eyes,” in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition*.
- [**Kelly 1970**] M. D. Kelly, 1970. Visual identification of people by computer. Tech. rep. AI-130, Stanford AI Project, Stanford, CA.
- [**Kervrann et autres 1997**] C. Kervrann, F. Davoine, P. Pérez, R. Forchheimer, and C. Labit, 1997. Generalized likelihood ratio-based face detection and extraction of mouth features, *Pattern Recog. Lett.* 18.
- [**Kim et autres 1998**] S.-H. Kim, N.-K. Kim, S. C. Ahn, and H.-G. Kim, 1998. Object oriented face detection using range and color information, in *IEEE Proc. of 3rd Int. Conf. on Automatic Face and Gesture Recognition*.
- [**Kim et Kim 2000**] S.-H. Kim and H.-G. Kim, 2000. Face detection using multi-modal information, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition*.
- [**Kirby et Sirovich 1990**] M. Kirby, and L. Sirovich, 1990. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. Patt. Anal. Mach. Intell.* 12.
- [**Kittler 1975**] J. Kittler, 1975. Mathematical methods of feature selection in pattern recognition. *International Journal of Man-Machine Studies*, 7:609–637.
- [**Kittler 1986**] J. Kittler, 1986. Feature selection and extraction. In T.Y. Young and K.S. Fu, eds, *Handbook of Pattern Recognition and Image Processing*, pp. 59-83. Academic Press, London.
- [**Koh et autres 1999**] L. H. Koh, S. Ranganath, M. W. Lee, and Y. V. Venkatesh, 1999. An integrated face detection and recognition system, in *Proceedings of the 10th International Conference on Image Analysis and Processing*.
- [**Kohonen 1995**] T. Kohonen, 1995. *Self-Organizing Maps*, Springer-Verlog, Berlin.
- [**Kotropoulos et Pitas 1997**] C. Kotropoulos and I. Pitas, 1997. Rule-based face detection in frontal views, in *Proc. Int. Conf. on Acoustic, Speech and Signal Processing*.
- [**Kruger et autres 1997**] N. Kruger, M. Potzsch, and C. V. D. Malsburg, 1997. Determination of face position and pose with a learned representation based on labelled graphs. *Image Vis. Comput.* 15, 665–673.
- [**Kumar et Poggio 2000**] V. Kumar and T. Poggio, 2000. Learning-based approach to real time tracking and analysis of faces, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition*.
- [**Kung et Taur 1995**] S. Y. Kung, and J. S. Taur, 1995. Decision-based neural networks with signal/image classification applications. *IEEE Trans. Neural Netw.* 6, 170–181.

- [**LaCascia et autres 2000**] M. LaCascia, S. Sclaroff, and V. Athitsos, 2000. Fast, reliable head tracking under varying illumination: An approach based on registration of textured-mapped 3D models, *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 322–336.
- [**Lades et autres 1993**] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. V. D. Malsburg, R. Wurtz, and W. Konen, 1993. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.* 42, 300–311.
- [**Lam et Yan 1994**] K. M. Lam and H. Yan, 1994. Facial feature location and extraction for computerised human face recognition, in *Int. Symposium on information Theory and Its Applications*, Sydney, Australia, Nov. 1994.
- [**Lanitis et autres 1995**] A. Lanitis, C. J. Taylor, and T. F. Cootes, 1995. Automatic face identification system using flexible appearance models. *Image Vis. Comput.* 13, 393–401.
- [**Lawrence et autres 1997**] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, 1997. Face recognition: A convolutional neural-network approach. *IEEE Trans. Neural Netw.* 8, 98–113.
- [**Lee et autres 1996**] C. H. Lee, J. S. Kim, and K. H. Park, 1996. Automatic human face location in a complex background, *Pattern Recog.* 29, 1877–1889.
- [**Lew et Huijismans 1996**] M. S. Lew and N. Huijismans, 1996. Information theory and face detection, in *Proc. of International Conference on Pattern Recognition*.
- [**Li et autres 2000**] Y. Li, A. Goshtasby, and O. Garcia, 2000. Detecting and tracking human faces in videos, in *Proceedings of the 15th International Conference on Pattern Recognition*, Vol. I, p. 2A.
- [**Li et Lu 1999**] S. Z. Li and J. Lu, 1999. Face recognition using the nearest feature line method. *IEEE Trans. Neural Netw.* 10, 439–443.
- [**Li et Roeder 1995**] X. Li and N. Roeder, 1995. Face contour extraction from front-view images, *Pattern Recog.* 28.
- [**Lin et autres 1997**] S. H. Lin, S. Y. Kung, and L. J. Lin, 1997. Face recognition/detection by probabilistic decision based neural network. *IEEE Trans. Neural Netw.* 8, 114–132.
- [**Liu et Wang 2000**] Z. Liu and Y. Wang, 2000. Face detection and tracking in video using dynamic programming, in *Proceedings of the 2000 International Conference on Image Processing*, pp. MA02.08.
- [**Liu et Wechsler 2000a**] C. Liu and H. Wechsler, 2000a. Evolutionary pursuit and its application to face recognition. *IEEE Trans. Patt. Anal. Mach. Intell.* 22, 570–582.
- [**Liu et Wechsler 2000b**] C. Liu and H. Wechsler, 2000b. Robust coding scheme for indexing and retrieval from large face databases. *IEEE Trans. Image Process.* 9, 132–137.
- [**Liu et Wechsler 2001**] C. Liu and H. Wechsler, 2001. A shape- and texture-based enhanced fisher classifier for face recognition. *IEEE Trans. Image Process.* 10, 598–608.
- [**Low 1998**] B. K. Low, 1998. *Computer Extraction of Human Faces*, PhD thesis, Dept. of Electronic and Electrical Engineering, De Montfort University.
- [**Low et Ibrahim 1997**] B. K. Low and M. K. Ibrahim, 1997. A fast and accurate algorithm for facial feature segmentation, in *Proceedings International Conference on Image Processing*.
- [**Luthon et Lievin 1997**] F. Luthon and M. Lievin, 1997. Lip motion automatic detection, in *Scandinavian Conference on Image Analysis*, Lappeenranta, Finland.
- [**Lv et autres 2000**] X.-G. Lv, J. Zhou, and C.-S. Zhang, 2000. A novel algorithm for rotated human face detection, in *IEEE Conference on Computer Vision and Pattern Recognition*.

- [**Manjunath et autres 1992**] B. S. Manjunath, R. Chellappa, and C. V. D. Malsburg, 1992. A feature based approach to face recognition. In Proceedings, IEEE Conference on Computer Vision and Pattern Recognition. 373–378.
- [**Marchand-Maillet et Merialdo 1999**] S. Marchand-Maillet and B. Merialdo, 1999. Pseudo two-dimensional hidden markov models for face detection in colour images, in Proceedings Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA).
- [**Mardia et autres 1979**] K. V. Mardia, J. T. Kent, and J. M. Bibby, 1979. Multivariate Analysis, Academic Press, San Diego.
- [**Marqués et Vilaplana 2000**] F. Marqués and V. Vilaplana, 2000. A morphological approach for segmentation and tracking of human faces, in Proceedings of the 15th International Conference on Pattern Recognition, Vol. I, P. 3A.
- [**Marr et Hildreth 1980**] D. Marr and E. Hildreth, 1980. Theory of edge detection, in Proc. of the Royal Society of London.
- [**Martinez 2002**] A. Martinez, 2002. Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class. IEEE Trans. Patt. Anal. Mach. Intell. 24, 748–763.
- [**Maurer et Malsburg 1996a**] T. Maurer and C. V. D. Malsburg, 1996a. Single view based recognition of faces rotated in depth. In Proceedings, International Workshop on Automatic Face and Gesture Recognition. 176–181.
- [**McKenna et autres 1996**] S. McKenna, S. Gong, and J. J. Collins, 1996. Face tracking and pose representation, in British Machine Vision Conference, Edinburgh, Scotland, Sept. 1996.
- [**McKenna et autres 1998**] S. J. McKenna, S. Gong, and Y. Raja, 1998. Modelling facial colour and identity with Gaussian mixtures, Pattern Recog. 31.
- [**Meng et autres 2000**] L. Meng, T. Nguyen, and D. A. Castañón, 2000. An image-based bayesian framework for face detection, in IEEE Conference on Computer Vision and Pattern Recognition.
- [**Meng et Nguyen 2000**] L. Meng and T. Nguyen, 2000. Two subspace methods to discriminate faces and clutters, in Proceedings of the 2000 International Conference on Image Processing, p. TA07.03.
- [**Menser et Brunig 1999**] B. Menser and M. Brunig, 1999. Segmentation of human faces in color images using connected operators, in Proceedings International Conference on Image Processing.
- [**Moghaddam et Pentland 1994**] B. Moghaddam and A. Pentland, 1994. Face recognition using view-based and modular eigenspaces, in Automatic Systems for the Identification of Humans, SPIE, Vol. 2277.
- [**Moghaddam et Pentland 1997**] B. Moghaddam and A. Pentland, 1997. Probabilistic visual learning for object representation, IEEE Trans. Pattern Anal. Mach. Intell. 19(1).
- [**Moon et Phillips 2001**] H. Moon and P. J. Phillips, 2001. Computational and performance aspects of PCA-based face recognition algorithms. Perception, 30, 301–321.
- [**Nefian et Hayes 1998**] A. V. Nefian and M. H. Hayes III, 1998. Hidden Markov models for face recognition. In Proceedings, International Conference on Acoustics, Speech and Signal Processing. 2721–2724.
- [**Ng et Gong 1999**] J. Ng and S. Gong, 1999. Performing multi-view face detection and pose estimation using a composite support vector machine across the view sphere, in Proc. IEEE International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems.

- [**Okada et autres 1998**] K. Okada, J. Steffans, T. Maurer, H. Hong, E. Elagin, H. Neven, and C. V. D. Malsburg, 1998. The Bochum/USC Face Recognition System and how it fared in the FERET Phase III Test. In *Face Recognition: From Theory to Applications*, H. Wechsler, P. J. Phillips, V. Bruce, F. F. Soulie, and T. S. Huang, Eds. Springer-Verlag, Berlin, Germany, 186–205.
- [**Oliver et autres 2000**] N. Oliver, A. Pentland, and F. Bérard, 2000. LAFTER: A real-time face and lips tracker with facial expression recognition, *Pattern Recog.* 33, 1369–1382.
- [**ORL Database**] <http://www.uk.research.att.com/facedatabase.html>
- [**Osuna et autres 1997**] E. Osuna, R. Freund, and F. Girosi, 1997. Training support vector machines: An application to face detection, in *IEEE Proc. of Int. Conf. on Computer Vision and Pattern Recognition*, 6.
- [**Pavlidis 1982**] T. Pavlidis, 1982. *Graphics and Image Processing*, Computer Science Press, Rockville, MD.
- [**Penev et Atick 1996**] P. Penev and J. Atick, 1996. Local feature analysis: A general statistical theory for object representation. *Netw.: Computat. Neural Syst.* 7, 477–500.
- [**Penev et Sirovich 2000**] P. Penev and L. Sirovich, 2000. The global dimensionality of face space. In *Proceedings, International Conference on Automatic Face and Gesture Recognition*.
- [**Pentland et autres 1994**] A. Pentland, B. Moghaddam, and T. Starner, 1994. View-based and modular eigenspaces for face recognition. In *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*.
- [**Perlibakas 2004**] V. Perlibakas, 2004. Distance measures for PCA-based face recognition, In *Pattern Recognition Letters* 25, 711-724.
- [**Phillips 1998**] P. J. Phillips, 1998. Support vector machines applied to face recognition. *Adv. Neural Inform. Process. Syst.* 11, 803–809.
- [**Phillips et autres 1998**] P. J. Phillips, R. M. McCabe, and R. Chellappa, 1998. Biometric image processing and recognition. In *Proceedings, European Signal Processing Conference*.
- [**Phillips et autres 2000**] P. J. Phillips, H. Moon, , S. Rizvi, and P. Rauss, 2000. The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. Patt. Anal. Mach. Intell.* 22.
- [**Press et autres 1992**] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, 1992. *Numerical Recipes. The Art of Scientific Computing*, 2nd edn., Cambridge University Press, Cambridge.
- [**Propp et Samal 1992**] M. Propp and A. Samal, 1992. Artificial neural network architecture for human face detection, *Intell. Eng. Systems Artificial Neural Networks* 2, 535–540.
- [**Qian et Huang 1997**] R. Qian and T. Huang, 1997. Object detection using hierarchical mrf and map estimation, in *IEEE Proc. of Int. Conf. on Computer Vision and Pattern Recognition*.
- [**Raducanu et Grana 2000**] B. Raducanu and M. Grana, 2000. Face localization based on the morphological multiscale fingerprint, in *Proceedings of the 15th International Conference on Pattern Recognition*, Vol. II, p. 4B.
- [**Rajagopalan et autres 1998**] A. N. Rajagopalan, K. S. Kumar, J. Karlekar, R. Manivasakan, M. M. Patil, U. B. Desai, P. G. Poonacha, and S. Chaudhuri, 1998. Finding faces in photographs, in *Proceedings of International Conference on Computer Vision*.

- [**Ramos et autres 1997**] M. U. Ramos Sanchez, J. Matas, and J. Kittler, 1997. Statistical chromaticity models for lip tracking with b-splines, in Int. Conf. on Audio- and Video-Based Biometric Person Authentication, Crans Montana, Switzerland.
- [**Reinders et autres 1996**] M. J. T. Reinders, R. W. C. Koch, and J. J. Gerbrands, 1996. Tracking facial features in image sequences using neural networks, in IEEE Proc. of 2nd Int. Conf. on Automatic Face and Gesture Recognition, Vermont, Oct. 1996, pp. 230–235.
- [**Roth 1999**] D. Roth, 1999. “The SNoW Learning Architecture,” Technical Report UIUCDCS-R-99-2102, UIUC Computer Science Department.
- [**Roth et autres 2000**] D. Roth, M.-H. Yang, and N. Ahuja, 2000. A SNoW-based face detector, in Advances in Neural Information Processing Systems 12 (NIPS 12), MIT Press, Cambridge, MA.
- [**Rowley et autres 1998**] H. A. Rowley, S. Baluja, and T. Kanade, 1998. Neural network based face detection. IEEE Trans. Patt. Anal. Mach. Intell. 20.
- [**Rowley et autres 1998b**] H. A. Rowley, S. Baluja, and T. Kanade, 1998b. Rotation invariant neural network-based face detection, in Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recognition, pp. 38–44.
- [**Ruderman 1994**] D. L. Ruderman, 1994. The statistics of natural images. *Netw.: Comput. Neural Syst.* 5, 598–605.
- [**Saber et Tekalp 1998**] E. Saber and A. M. Tekalp, 1998. Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions, *Pattern Recog. Lett.* 19.
- [**Sakai et autres 1972**] T. Sakai, M. Nagao, and T. Kanade, 1972. Computer analysis and classification of photographs of human faces, in Proc. First USA—Japan Computer Conference, p. 2.7.
- [**Samal et Iyengar 1992**] A. Samal and P. Iyengar, 1992. Automatic recognition and analysis of human faces and facial expressions: A survey. *Patt. Recog.* 25, 65–77.
- [**Samal et Iyengar 1995**] A. Samal and P. A. Iyengar, 1995. Human face detection using silhouettes, *Int. J. Pattern Recog. Artificial Intell.* 9(6).
- [**Samaria 1994**] F. Samaria, 1994. Face recognition using hidden markov models. Ph.D. dissertation. University of Cambridge, Cambridge, U.K.
- [**Samaria et Young 1994**] F. Samaria and S. Young, 1994. HMM based architecture for face identification. *Image Vis. Comput.* 12, 537–583.
- [**Satoh 2000**] S. Satoh, 2000. Comparative evaluation of face sequence matching for content-based video access, in Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition.
- [**Schneiderman et Kanade 1998**] H. Schneiderman and T. Kanade, 1998. Probabilistic modeling of local appearance and spatial relationships for object recognition, in IEEE Conference on Computer Vision and Pattern Recognition, 6.
- [**Schneiderman et Kanade 2000**] H. Schneiderman and T. Kanade, 2000. A statistical model for 3D object detection applied to faces and cars, in IEEE Conference on Computer Vision and Pattern Recognition.
- [**Schneiderman et Kanade 2000a**] H. Schneiderman and T. Kanade, 2000a. Probabilistic modelling of local Appearance and spatial relationships for object recognition. In Proceedings, IEEE Conference on Computer Vision and Pattern Recognition. 746–751.
- [**Senior 1999**] A. W. Senior, 1999. Face and feature finding for a face recognition system, in Proceedings Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA).

- [**Shashua 1994**] A. Shashua, 1994. Geometry and photometry in 3D visual recognition. Ph.D. dissertation. Massachusetts Institute of Technology, Cambridge, MA.
- [**Shpungin et Movellan 2000**] B. E. Shpungin and J. R. Movellan, 2000. A Multi-threaded Approach to Real Time Face Tracking, Technical Report TR 2000.02, UCSD MPLab.
- [**Sirovich et Kirby 1987**] L. Sirovich and M. Kirby, 1987. Low-dimensional procedure for the characterization of human face. *J. Opt. Soc. Am.* 4, 519–524.
- [**Sobottka et Pitas 1996**] K. Sobottka and I. Pitas, 1996. Extraction of facial regions and features using color and shape information, in *Proc. of Int. Conf. on Pattern Recognition*.
- [**Song et Robinson 2000**] Q. Song and J. Robinson, 2000. A feature space for face image processing, in *Proceedings of the 15th International Conference on Pattern Recognition, Vol II*, p. 2B.
- [**Stewart 1973**] G.W. Stewart, 1973. *Introduction to Matrix Computation*. Academic Press, Orlando, FL.
- [**Sun et autres 1998**] Q. B. Sun, W. M. Huang, and J. K. Wu, 1998. Face detection based on color and local symmetry information, in *IEEE Proc. of 3rd Int. Conf. on Automatic Face and Gesture Recognition*.
- [**Sung et Poggio 1997**] K. Sung and T. Poggio, 1997. Example-based learning for view-based human face detection. *IEEE Trans. Patt. Anal. Mach. Intell.* 20, 39–51.
- [**Sung et Poggio 1998**] K.-K. Sung and T. Poggio, 1998. Example-based learning for view-based human face detection, *IEEE Trans. Pattern Anal. Mach. Intelligence* 20, 39–51.
- [**Swets et Weng 1996**] D. L. Swets and J. Weng, 1996. Discriminant analysis and eigenspace partition tree for face and object recognition from views. In *Proceedings, International Conference on Automatic Face and Gesture Recognition*. 192–197.
- [**Swets et Weng 1996b**] D. L. Swets and J. Weng, 1996b. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Patt. Anal. Mach. Intell.* 18, 831–836.
- [**Takacs et Wechsler 1997**] B. Takacs and H. Wechsler, 1997. Detection of faces and facial landmarks using iconic filter banks, *Pattern Recog.* 30.
- [**Tanaka et autres 1998**] M. Tanaka, K. Hotta, T. Kurita, and T. Mishima, 1998. Dynamic attention map by using model for human face detection, in *Proc. of International Conference on Pattern Recognition*.
- [**Terrillon et autres 1998**] J.-C. Terrillon, M. David, and S. Akamatsu, 1998. Automatic detection of human faces in natural scene images by use of a skin color model and of invariant moments, in *IEEE Proc. of 3rd Int. Conf. on Automatic Face and Gesture Recognition*.
- [**Terrillon et autres 2000**] J.-C. Terrillon, M. Shirazi, H. Fukamachi, and S. Akamatsu, 2000. Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition*.
- [**Terrillon et autres 2000b**] J. Terrillon, M. Shirazi, M. Sadek, H. Fukamachi, and S. Akamatsu, 2000b. Invariant face detection with support vector machines, in *Proceedings of the 15th International Conference on Pattern Recognition, Vol. IV*, p. 4B.
- [**Thisted 1988**] R.A. Thisted, 1988. *Elements of Statistical Computing. Numerical Computation*. Chapman & Hall, New York.

- [**Tsapatsoulis et autres 2000**] N. Tsapatsoulis, Y. Avrithis, and S. Kollias, 2000. Efficient face detection for multimedia applications, in Proceedings of the 2000 International Conference on Image Processing, p. TA07.11.
- [**Turk et Pentland 1991**] M. Turk and A. Pentland, 1991. Eigenfaces for recognition. *J. Cogn. Neurosci.* 3, 71–86.
- [**Valentin et autres 1994**] D. Valentin, H. Abdi, A. J. O’Toole, and G. Cottrell, 1994. Connectionist models of face processing: A survey, *Pattern Recog.* 27, 1209–1230.
- [**Van Beek et autres 1992**] P. J. L. Van Beek, M. J. T. Reinders, B. Sankur, and J. C. A. Van Der Lubbe, 1992. Semantic segmentation of videophone image sequences, in Proc. of SPIE Int. Conf. on Visual Communications and Image Processing, pp. 1182–1193.
- [**Vapnik 1995**] V. N. Vapnik, 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, NY.
- [**Viennet et Fogelman Soulié 1998**] E. Viennet and F. Fogelman Soulié, 1998. Connectionist methods for human face processing, in *Face Recognition: From Theory to Application*. Springer-Verlag, Berlin/New York.
- [**Vincent et autres 1992**] J. M. Vincent, J. B. Waite, and D. J. Myers, 1992. Automatic location of visual features by a system of multilayered perceptrons, in *IEE Proceedings-F*, Dec. 1992, Vol. 139.
- [**Viola et Jones 2001**] P. Viola and M. Jones, 2001. Rapid object detection using a boosted cascade of simple features. In Proceedings, IEEE Conference on Computer Vision and Pattern Recognition.
- [**Wang et Tan 2000**] J. Wang and T. Tan, 2000. A new face detection method based on shape information, *Pattern Recog. Lett.* 21, 463–471.
- [**Wang et Chang 1997**] H. Wang and S.-F. Chang, 1997. A highly efficient system for automatic face region detection in MPEG video, *IEEE Trans. Circuits Systems Video Technol.* 7, 615–628.
- [**Wang et Sung 1999**] J. G. Wang and E. Sung, 1999. Frontalview face detection and facial feature extraction using color and morphological operations, *Pattern Recog. Lett.* 20, 1053–1068.
- [**Weber et Herrera Hernandez 1999**] F. Weber and A. Herrera Hernandez, 1999. Face location by template matching with a quadratic discriminant function, in Proc. IEEE International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems.
- [**Wei et Sethi 1999**] G. Wei and I. K. Sethi, 1999. Face detection for image annotation, *Pattern Recog. Lett.* 20, 1313–1321.
- [**Wiskott et autres 1997**] L. Wiskott, J.-M. Fellous, and C. Von Der Malsburg, 1997. Face recognition by elastic bunch graph matching. *IEEE Trans. Patt. Anal. Mach. Intell.* 19, 775–779.
- [**Wong et autres 1995**] C. Wong, D. Kortenkamp, and M. Speich, 1995. A mobile robot that recognises people, in *IEEE Int. Conf. on Tools with Artificial Intelligence*.
- [**Woo et Oh 1999**] T. W. Yoo and I. S. Oh, 1999. A fast algorithm for tracking human faces based on chromatic histograms, *Pattern Recog. Lett.* 20, 967–978.
- [**Wu et autres 2000**] H. Wu, Q. Chen, and M. Yachida, 2000. Face detection from color images using a fuzzy pattern matching method, *IEEE Trans. Pattern Anal. Mach. Intell.* 21, 557–563.
- [**Wu et Zhou 2003**] J. Wu and Z.-H. Zhou, 2003. Efficient face candidates selector for face detection, *Pattern Recognition* 36 (2003), 1175–1186.
- [**Wyszecki et Stiles 1967**] G. Wyszecki and W. S. Stiles, 1967. *Color Science*, Wiley, New York.

- [**Yachi et autres 2000**] K. Yachi, T. Wada, and T. Matsuyama, 2000. Human head tracking using adaptive appearance models with a fixed-viewpoint pan-tilt-zoom camera, in Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition.
- [**Yang 2000**] L. Yang, 2000. Multiple-face tracking system for general region-of-interest video coding, in Proceedings of the 2000 International Conference on Image Processing, p. MA09.13.
- [**Yang and Huang 1994**] G. Yang and T. S. Huang, 1994. Human face detection in a complex background, *Pattern Recog.* 27, 53–63.
- [**Yang et autres 2000**] M.-H. Yang, N. Ahuja, and D. Kriegman, 2000. Face detection using mixtures of linear subspaces, in Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition.
- [**Yang et autres 2002**] M. H. Yang, D. Kriegman, and N. Ahuja, 2002. Detecting faces in images: A survey. *IEEE Trans. Patt. Anal. Mach. Intell.* 24, 34–58.
- [**Yang et Waibel 1996**] J. Yang and A. Waibel, 1996. A real-time face tracker, in *IEEE Proc. of the 3rd Workshop on Applications of Computer Vision*, Florida.
- [**Yuille et autres 1992**] A. L. Yuille, D. S. Cohen, and P. W. Hallinan, 1992. Feature extraction from faces using deformable templates. *Int. J. Comput. Vis.* 8, 99–112.
- [**Zhao 1999**] W. Zhao, 1999. Robust Image Based 3D Face Recognition, Ph.D. dissertation. University of Maryland, College Park, MD.
- [**Zhao et autres 1998**] W. Zhao, R. Chellappa, and A. Krishnaswamy, 1998. Discriminant analysis of principal components for face recognition. In Proceedings, International Conference on Automatic Face and Gesture Recognition. 336–341.
- [**Zhao et autres 1999**] W. Zhao, R. Chellappa, and P. J. Phillips, 1999. Subspace linear discriminant analysis for face recognition. Tech. rep. CAR-TR-914, Center for Automation Research, University of Maryland, College Park, MD.
- [**Zhao et autres 2003**] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, 2003. Face Recognition: A Literature Survey. *ACM Computing Surveys*, Vol. 35, No 4, Dec. 2003, pp. 399-458.
- [**Zhu et autres 2000**] Y. Zhu, S. Schwartz, and M. Orchard, 2000. Fast face detection using subspace discriminant wavelet features, in *IEEE Conference on Computer Vision and Pattern Recognition*.

## **Résumé:**

La détection et la reconnaissance des visages humains dans des images trouvent de nombreuses applications potentielles telles que la surveillance, les systèmes de recensement et les interfaces homme machine, elles constituent des tâches très difficiles à cause des variations importantes dans l'apparence des visages humains.

Dans ce mémoire, on propose deux méthodes pour la détection de visages frontaux dans un environnement complexe. La première extrait en premier lieu les yeux qui sont les parties les plus sombres du visage humain, elle vérifie par la suite les candidats possibles de visages par une classification dans l'espace propre de ses formes en visage ou non visage. Dans la deuxième méthode, on a développé un modèle de segmentation dans l'espace des couleurs normalisées rgb, on a ajouté des contraintes géométriques pour améliorer le résultat de la segmentation et on a vérifié la présence de visages par la recherche des segments analogues aux yeux.

Pour la reconnaissance de visage, on a implémenté la technique des visages propres avec deux bases de données (CVL et ORL), on a comparé les performances relatives aux différentes mesures de similitude. Enfin, on a implémenté aussi la méthode des visages propres des images réduites qui donne de meilleurs résultats.

## **Abstract:**

Face detection and recognition in images have a wide range of applications, such as supervision, census systems, and human-machine interfaces; they are difficult because of the huge variations in the appearance of human faces.

In this thesis, we propose two methods for frontal face detection in cluttered backgrounds. The first method extracts the eyes which are the darkest regions in the human face, and then it verifies the possible face candidates by a classification of the face patterns to face or non-face in the face space. In the second method, we have developed a segmentation model in the rgb color space, we have added some geometrical constraints to improve the segmentation result, and then we have verified the face presence by the finding of eye-analogue segments.

For the face recognition purpose, we have implemented the eigenfaces method with two face databases (CVL and ORL), we have compared the performances related to the use of different similarity measures. Finally, we have implemented the reduced image eigenfaces method which provides better results.

## **ملخص:**

التأكد من وجود الوجه و التعرف عليه في الصور تعرف تطبيقات عديدة و مهمة مثل المراقبة، أجهزة الإحصاء و الواجهات بين الإنسان و الآلة، و هي تمثل مهمات جد صعبة بسبب التغيرات المهمة في مظهر الوجه. في هذه المذكرة، اقترحنا طريقتين جديدتين للتأكد من وجود الوجه في بيئة معقدة. الأولى تستخرج العينين الذين يشكلان المناطق الداكنة من الوجه الإنساني، ثم تتحقق من وجود الأوجه بواسطة تصنيف في الفضاء لهذه الأنماط إلى وجوه و غيرها. بالنسبة للطريقة الثانية، لقد أنشأنا وسيلة للتقسيم في فضاء الألوان المعدلة، ثم أضفنا متطلبات هندسية لتحسين النتيجة و تحققنا من وجود الوجه بالبحث عن المناطق الشبيهة بالعينين. في ما يخص التعرف على الوجه، استخدمنا طريقة الأوجه الذاتية مع قاعدتي البيانات CVL و ORL و قارنا النتائج بالنسبة لعدة مقاييس للتشابه. أخيراً، لقد استخدمنا طريقة الأوجه الذاتية المختزلة التي أعطت نتائج أفضل.