



UNIVERSITE SULTAN MOULAY SLIMANE
Faculté des Sciences et Techniques
Béni-Mellal



Centre d'Etudes Doctorales : Sciences et Techniques
Formation Doctorale : Mathématiques et Physique Appliquées

THÈSE

Présentée par

Abdellatif HAJRAOUI

Pour l'obtention du grade de

Docteur

Spécialité : Informatique

Contribution à la reconnaissance de visages humains

Soutenue le 21 Novembre 2015 devant la commission d'examen :

Pr. Mohamed SAJIEDDINE	Professeur à la FST- Béni Mellal	Président
Pr. Abdelouhab ZEROUAL	Professeur à la FSS- Marrakech	Rapporteur
Pr. Moha Mrabet HASSANI	Professeur à la FSS- Marrakech	Rapporteur
Pr. Hicham AISSAOUI	Professeur à la FST- Béni Mellal	Rapporteur
Pr. Mohamed SABRI	Professeur à la FST- Béni Mellal	Directeur de thèse
Pr. Mohamed FAKIR	Professeur à la FST- Béni Mellal	Co-directeur de thèse

Je dédie cette thèse

à Siham, Louay et Raghad pour leur amour inconditionnel,

à ma famille qui m'a soutenu,

et à mes amis proches qui m'ont encouragé...

Remerciements

Pour réaliser ce manuscrit et le travail qu'il présente, j'ai largement bénéficié de l'aide de nombreuses personnes. Je tiens à les remercier très sincèrement.

*Je voudrai tout d'abord témoigner mes sincères remerciements et ma profonde reconnaissance au professeur **Mohamed Sabri**, mon directeur de thèse. Travailler avec vous a été enrichissant, merci de m'avoir encouragé durant toute ma thèse. Merci pour la confiance que tu m'as témoignée. Merci pour ton aide précieuse sur le plan scientifique et administratif. Merci pour ta patience et ta disponibilité. Je te serai toujours reconnaissant pour l'effort que tu as investi dans cette thèse.*

*Mes chaleureux remerciements s'adressent également à mon co-directeur de thèse, le professeur **Mohamed Fakir**, pour sa disponibilité, pour son attention de tout instant sur mes travaux et ses conseils tout au long de cette thèse. Sa confiance et son encouragement ont été un élément moteur pour moi.*

*Ensuite, j'aimerais sincèrement remercier les membres du jury, les professeurs **Mohamed Sajieddine**, **Abdelouhab Zeroual**, **Moha Mrabet Hassani** et **Hicham Aissaoui** pour le temps précieux qu'ils ont accepté d'accorder à la lecture de cette thèse.*

D'entrée de jeu, j'aimerais exprimer ma sincère reconnaissance aux professeurs et au personnel du CED.

Merci à ma femme et à mes deux enfants pour leur amour et leur soutien lors de moments difficiles.

Enfin, mes vifs remerciements s'adressent à toute ma famille et à mes amis proches pour leur encouragement et leur appui constant.

Résumé

Bien que la reconnaissance de visages humains soit un domaine difficile à cause de la multitude des challenges à surmonter, il est très important de s'y intéresser et d'y investir vu la multiplicité et la variété de ses champs d'application (vérification d'identité, contrôle d'accès physique et logique, vidéosurveillance, interfaces homme-machine ...). Le travail réalisé au cours de cette thèse s'inscrit dans ce contexte dans l'objectif de la mise en œuvre d'une architecture logicielle complète d'un système automatisé de reconnaissance de visages. Afin de bien cerner les problèmes qui peuvent dégrader les performances d'un tel système, nous avons effectué plusieurs études à différents niveaux de sa chaîne de traitement et nous avons proposé par la suite des solutions. Nous avons développé dans un premier temps une nouvelle approche destinée à la détection et la localisation des visages dans une image 2D couleur. Cette approche est fondée sur trois points pivots : la couleur de la peau, la ligne de partage des eaux (LPE) et les ondelettes de Gabor. Dans un second temps, nous nous sommes focalisés sur l'extraction des vecteurs caractéristiques des visages (signatures) à l'aide d'une approche originale qui s'est étalée en deux versions, la première est basée sur l'ACP et l'ADL, tandis que la deuxième est élaborée dans le but de palier les inconvénients de la première en intégrant le descripteur de Gabor et un estimateur de pose. Finalement, pour valider ce travail de recherche, nous avons testé ces approches sur les bases de données standards: Caltech_Web, AT&T et FERET.

Mots Clefs : Biométrie, reconnaissance de visages, détection de visages, ACP, ADL, descripteur de Gabor, estimateur de pose, réseaux de neurones, machines à vecteurs de support (SVM), distances métriques...

Abstract

Although recognition of human faces is a difficult area because of the multitude of challenges to overcome, it is very important to be interested and invest given the multiplicity and variety of its fields of application (verification of identity, physical and logical access control, video-surveillance, human machine interfaces, ...). The work carried over this thesis is part of this context with the objective of implementing a complete software architecture of an automated face recognition. Order to properly identify the problems that can degrade the performance of such a system, we performed several studies at various levels of her processing chain and we proposed thereafter solutions. We have developed in the first instance a new approach for the detection and location of faces in a 2D color image. This approach was based on three pivotal points: the color of the skin, the watershed line and the Gabor wavelets. Secondly, we focused on the extraction of feature vectors faces (signatures) using an original approach that was spread in two versions, the first is based on the PCA and the LDA, while the second is made in the aim bearing of the disadvantages of the first by integrating the descriptor of Gabor and a pose estimator. Finally, to validate this research work, we tested these approaches on standard databases: Caltech_Web, AT & T and FERET.

Key words: Biometrics, face recognition, face detection, PCA, LDA, descriptor of Gabor, pose estimator, neural networks, Supports Vectors Machines (SVM), metric distances...

Table des matières

Table des matières	4
Liste des figures	9
Liste des tableaux	12
Liste des abréviations	13
Chapitre 1. Introduction générale	16
1.1. Contexte et motivations de la thèse	16
1.2. Objectifs de la thèse	21
1.3. Principales contributions	22
1.4. Organisation de la thèse	23
Chapitre 2. La reconnaissance de visages	27
2.1. Introduction	27
2.2. Applications de la reconnaissance de visages	27
2.3. Modes de fonctionnement de la reconnaissance de visages	28
2.4. Architecture de base d'un système de reconnaissance de visages	28
2.5. Challenges liés à la reconnaissance de visages	33
2.5.1. Facteurs extrinsèques	33
2.5.2. Facteurs intrinsèques	35
2.6. Outils d'évaluation d'un système de reconnaissance de visages	36
2.6.1. Le mode de vérification	36
2.6.2. Le mode d'identification	38
2.7. Bases de données des images de visage	40
2.7.1. La base de données FERET	40
2.7.2. La base AT&T	40
2.7.3. La Base AR	41
2.7.4. La Base Extended Yale B (B+)	41
2.7.5. La base Multi-PIE Face Database	41
2.8. Conclusion	42

Chapitre 3. Etat de l'art en reconnaissance de visages	44
3.1. Introduction.....	44
3.2. Méthodes globales	46
3.2.1. La méthode Eigenfaces ou ACP	46
3.2.2. Variantes de la méthode ACP	47
3.2.3. La méthode Fisherfaces ou ADL de Fisher	49
3.2.4. Variantes de la méthode ADL de Fisher.....	49
3.2.5. Analyse en Composantes Indépendantes (ACI)	50
3.2.6. Transformée de cosinus discrète (DCT)	51
3.3. Méthodes locales	52
3.3.1. Méthodes basées sur des caractéristiques locales.....	52
3.3.2. Méthodes basées sur l'apparence locale du visage.....	56
3.4. Méthodes hybrides.....	57
3.5. Comparaison	58
3.6. Conclusion	59
Chapitre 4. Détection et localisation de visages	61
4.1. Introduction.....	61
4.2. Méthodes de détection de visages.....	62
4.2.1. Approches basées sur les connaissances acquises	62
4.2.2. Approches basées sur le Template-matching.....	62
4.2.3. Approches basées sur l'apparence	63
4.2.4. Approches basées sur des caractéristiques invariantes	64
4.3. Principe de base de l'approche proposée.....	65
4.4. Détecteur de peau	67
4.4.1. Espaces de couleurs	67
4.4.2. Modèles de la distribution de la couleur de la peau.....	68
4.4.3. Le modèle de la couleur de peau proposé.....	68
4.5. Segmentation en régions de peau connexes et significatives	71
4.5.1. Présentation de la ligne de partage des eaux (LPE).....	71
4.5.2. Construction de la ligne de partage des eaux (LPE).....	73
4.5.3. L'algorithme de la LPE adopté	74

4.5.4. Prétraitement avant la segmentation de la LPE	74
4.5.5. Localisation des régions de peau connexes significatives	75
4.6. Classification visage et non-visage.....	76
4.6.1. Présentation du Filtre de Gabor (ondelette de Gabor)	77
4.6.2. Filtres de Gabor en modélisation faciale	79
4.6.3. Filtres de Gabor dans notre approche de détection de visage.....	80
4.6.4. Modèle de visage	82
4.6.5. Organigramme de décision visage ou non visage	83
4.7. Evaluation de la détection et la localisation des visages	84
4.8. Normalisation	85
4.8.1. Normalisation géométrique.....	85
4.8.2. Normalisation photométrique	86
4.9. Conclusion	86
Chapitre 5. Extraction des caractéristiques.....	88
5.1. Introduction.....	88
5.2. Première approche d'extraction: ACP-ADL double	89
5.2.1. Principe de base de l'approche	89
5.2.2. Description détaillée de l'approche	89
5.2.3. Discussion	94
5.3. Deuxième approche d'extraction des caractéristiques : Descripteur de Gabor, Estimateur de pose et ACP-ADL double.....	95
5.3.1. Principe de base de l'approche	95
5.3.2. Descripteur de Gabor	96
5.3.3. Estimateur de pose	97
5.3.4. Description détaillée de la deuxième approche d'extraction.....	100
5.4. Conclusion	102
Chapitre 6. Classification et décision.....	105
6.1. Introduction.....	105
6.2. Distances métriques	106
6.2.1. Définition d'une distance métrique.....	106
6.2.2. Les distances Euclidiennes	106
6.2.3. La distance de Mahalanobis.....	107

6.2.4. Distances métriques en reconnaissance de visages.....	108
6.3. Les réseaux de neurones	109
6.3.1. Définitions.....	109
6.3.2. Le neurone formel.....	109
6.3.3. Applications des réseaux de neurones	111
6.3.4. Architectures des réseaux de neurones	112
6.3.5. Les perceptrons multicouches.....	114
6.3.6. Réseaux de neurones en reconnaissance de visages	115
6.4. Les séparateurs à vastes marges (SVM).....	116
6.4.1. Présentation.....	116
6.4.2. Fondements de base des SVM	116
6.4.3. SVM à plusieurs classes.....	118
6.4.4. SVM en reconnaissance de visages	120
6.5. Conclusion	121
Chapitre 7. Résultats, évaluations et discussions	123
7.1. Introduction.....	123
7.2. Détection et localisation de visages.....	123
7.2.1. La base Caltech_10K_WebFaces	123
7.2.2. Evaluation du détecteur de pixels de peau.....	124
7.2.3. Evaluation de l'approche de détection et de localisation de visages	127
7.3. Extraction des caractéristiques et classification	129
7.3.1. La base AT&T	130
7.3.2. La base Color FERET.....	131
7.3.3. Protocole expérimental 1	132
7.3.4. Protocole expérimental 2	138
7.3.5. Protocole expérimental 3	138
7.3.6. Protocole expérimental 4	139
7.3.7. Protocole expérimental 5	141
7.4. Conclusion	142
Bilan et Perspectives.....	144
Conclusion générale.....	144

Perspectives.....	146
Annexe A. Espaces de couleurs	148
Annexe B. Algorithme de rétro-propagation du gradient	151
Diffusion scientifique de la recherche	154
Articles de revues	154
Communications de conférences internationales	154
Bibliographie	155

Liste des figures

Fig. 1.1. Principales modalités biométriques.....	18
Fig. 1.2. Analyse Zephyr (source: IBG)	19
Fig. 1.3. Statistiques des revenus des technologies biométriques 2009-2014	20
Fig. 2.1. Architecture d'un système de reconnaissance de visages	30
Fig. 2.2. Exemples de visages avec variation d'illumination.....	34
Fig. 2.3. Exemples de visages avec variation de pose	35
Fig. 2.4. Exemples de visages avec des occultations partielles.....	35
Fig. 2.5. Exemples de visages avec variation de l'expression faciale.....	36
Fig. 2.6. Illustration des erreurs : Faux Rejet et Fausse Acceptation.....	37
Fig. 2.7. Détermination du point d'équivalence des erreurs (EER)	38
Fig. 2.8. Illustration de la courbe ROC.....	38
Fig. 2.9. Exemples de courbe CMC de la méthode ACP	39
Fig. 2.10. Extrait de la base de données de visage AR.....	41
Fig. 2.11. Extrait de la base de données de visage Extended Yale B (B+)	41
Fig. 2.12. Extrait de la base de données de visage Multi-PIE	42
Fig. 3.1. Trois types de représentation d'une image de visage.....	45
Fig. 3.2. Les 17 premiers visages propres (eigenfaces) de la base AT&T	47
Fig. 3.3. Exemples des images de base obtenues avec la méthode ACI.	51
Fig. 3.4. La répartition des points détectés et les mesures choisies dans trois approches géométriques.	53
Fig. 3.5. Exemple de graphe élastique avec une topologie de grille	54
Fig. 3.6. Exemple de graphe élastique avec des points caractéristiques	55
Fig. 3.7. Exemple de représentation d'un visage par la méthode d'apparence local basée sur les histogrammes LBP.	57
Fig. 4.1. Exemples de caractéristiques pseudo-Haar utilisées par Viola et Jones.....	64
Fig. 4.2. Schéma synoptique de l'approche proposée pour la détection et la localisation de visage.....	66
Fig. 4.3. Graphe d'induction du modèle de la couleur de peau proposé	70
Fig. 4.4. Illustration de l'étape de la détection des pixels de peau.	71

Fig. 4.5. Illustration de la LPE topographique	72
Fig. 4.6. Exemple de segmentation d'une image en niveaux de gris par la LPE	72
Fig. 4.7. Illustration du chemin géodésique entre les deux points P1 et P2.	73
Fig. 4.8. Exemple de SKIZ géodésique	74
Fig. 4.9. Illustration de la segmentation par la LPE d'une image binaire.	75
Fig. 4.10. Illustration des étapes de masque et de localisation des régions de peau	76
Fig. 4.11. La partie réelle d'un exemple de filtre de Gabor.	78
Fig. 4.12. La représentation de la partie réelle de 4 filtres de Gabor ayant respectivement l'orientation θ : 0 ; $\pi/4$; $\pi/2$; $3\pi/4$	78
Fig. 4.13. La représentation de la partie réelle de 4 filtres de Gabor ayant respectivement la fréquence centrale f_0 : $1/8$; $1/12$; $1/14$; $1/16$	79
Fig. 4.14. Représentation des bandes passantes fréquentielles (Δf) et directionnelles ($\Delta\Omega$) d'un filtre de Gabor.	79
Fig. 4.15. Résultat de filtrage d'une image de visage qui a un angle d'inclinaison $+21.64^\circ$ par des filtres ayant la même fréquence centrale et différentes valeurs d'orientation	81
Fig. 4.16. Résultat de filtrage d'une image de visage avec un angle d'inclinaison $+21.64^\circ$ par des filtres ayant la même valeur de θ_v et différentes valeurs de la fréquence centrale	81
Fig. 4.17. Modèle de visage.....	83
Fig. 4.18. Organigramme de décision visage ou non visage	84
Fig. 4.19. Exemple d'opération d'égalisation d'histogramme : en haut l'image en niveau de gris avant égalisation et en bas après égalisation	86
Fig. 5.1. Schéma synoptique de la première approche d'extraction des caractéristiques	90
Fig. 5.2. Deux exemples montrant le manque de correspondance en raison de la perte d'information due à la forte variation d'illumination (en haut) ou la variation de pose (en bas)	95
Fig. 5.3. Processus d'extraction du vecteur caractéristique par le descripteur de Gabor	97
Fig. 5.4. Illustration de quelques repères de références utilisés dans l'estimation de pose : Repère corps, angles de rotation et repère caméra.....	98
Fig. 5.5. Paramétrage de la pose du visage par rapport au plan frontal du visage comme référence.....	98
Fig. 5.6. Exemple d'illustration du principe de fonctionnement de l'estimateur de pose proposé.....	100
Fig. 5.7. Schéma synoptique de la deuxième approche d'extraction	102

Fig. 6.1. Exemple d'architecture d'un réseau de neurone montrant les éléments de base qui le constituent	110
Fig. 6.2. Illustration de l'analogie entre un neurone biologique (à gauche) et un neurone formel (à droite)	111
Fig. 6.3. Exemple de perceptron multicouche à une seule couche cachée	115
Fig. 6.4. Hyperplans séparateurs arbitraires	117
Fig. 6.5. Illustration des notions: hyperplan optimal, vecteurs de support et marge maximale	118
Fig. 6.6. Exemple de transformation de l'espace : (a) Cas non linéairement séparable dans un espace à deux dimensions et (b) Cas linéairement séparable dans un espace à trois dimensions	119
Fig. 7.1. Détection des pixels de la peau avec différents modèles	125
Fig. 7.2. Deux exemples de détection des pixels de la peau avec notre modèle de peau proposé	126
Fig. 7.3. Exemple de détection de visages (a) notre approche et (b) algorithme de Viola-Jones	128
Fig. 7.4. Extrait de la base AT&T	131
Fig. 7.5. Des exemples extraits de la base Color FERET	132
Fig. 7.6. Les courbes CMC de la méthode ACP en utilisant dans la classification l'une des distances : city-block, Euclidienne et Mahalanobis	133
Fig. 7.7. Les courbes CMC de la méthode ACP-ADL double en utilisant dans la classification l'une des distances : city-block, Euclidienne et Mahalanobis	134
Fig. 7.8. Les courbes CMC de la méthode ACP-ADL double en utilisant dans la classification PMC ou SVM	134
Fig. 7.9. Les courbes CMC de la méthode Gabor + ACP-ADL double en utilisant dans la classification l'une des distances : city-block, Euclidienne et Mahalanobis	135
Fig. 7.10. Histogramme représentant les taux de reconnaissance des méthodes ACP, ACP-ADL double et Gabor + ACP-ADL double dans la base AT&T	136
Fig. 7.11. Illustration des projections des images de visage de 3 personnes	137
Fig. 7.12. Histogramme représentant les taux de reconnaissance des méthodes ACP, ACP-ADL double et Gabor + ACP-ADL double dans la base Color FERET	140
Fig. A.1. Cube de Maxwell	148
Fig. B. 1. Perceptron multicouche à une seule couche cachée.	152

Liste des tableaux

Tableau 7.1. Les taux RAR et FAR obtenus par notre approche de détection et l'algorithme de Viola-Jones.....	127
Tableau 7.2. Les temps de détection des deux méthodes : notre approche de détection et l'algorithme de Viola-Jones.....	128
Tableau 7.3. Taux de reconnaissance des trois méthodes ACP, ACP-ADL double et Gabor + ACP-ADL double.....	135
Tableau 7.4. Temps moyen d'identification en (s) des trois méthodes ACP, ACP-ADL double et Gabor + ACP-ADL double	135
Tableau 7.5. Taux de reconnaissance des méthodes ACP, ACP-ADL double et Gabor + ACP-ADL double avec différentes variations de l'apparence du visage.	139
Tableau 7.6. La classification théorique des valeurs de pose de test.....	141

Liste des abréviations

Terme	Signification
2D	Représentation à 2 Dimensions d'un visage
3D	Représentation à 3 Dimensions d'un visage
2.5D	Image de profondeur d'un visage
ACI	Analyse en Composantes I ndépendantes
ACP	Analyse en Composantes P incipales
ACP2D	Analyse en Composantes P incipales Bidimensionnelle
ADL	Analyse D iscriminante L inéaire
ARL	Army R esearch L aboratory
ART	Adaptative R esonance T heorie
CCD	Charge Coupled D evises
CMC	Cumulative M atch Characteristic
CMOS	Complementarity M etal O xide Semiconductor
CVC	Computer Vision C enter
DARPA	Defense A dvanced R esearch P rojects A gency
DCT	Discrete Cosine T ransform
DLA	Dynamic L ink A rchitecture
EBGM	Elastic B unch G raph M atching
EER	Equal E rror R ate
EGM	Elastic G raph M atching
FAR	False A cceptance R ate
FERET	Face R Ecognition T echnologie
Fig	F igure
FMR	False M atch R ate
FOCS	Face and O culaire C hallenge S eries
FRR	False R ejection R ate

FRVT	Face Recognition Vendor Test
HOG	Histogrammes des Orientations de Gradients
HOS	High Order Statistics
HSV	Hue Saturation Value
IBG	International Biometric Group
ICA	Independent Component Analysis
KDDA	Kernel Direct Discriminant Analysis
KFDA	Kernel Fisher Discriminant Analysis
KICA	Kernel Independant Component Analysis
KPCA	Kernel Principal Component Analysis
LBP	Local Binary Pattern
LCA	Local Component Analysis
LDA	Linear Discriminant Analysis
LPE	Ligne de Partage des Eaux
MBE	Multiple Biometric Evaluation
MBGC	Multiple Biometric Grand Challenge
MEGM	Morphological Elastic Graph Matching
MLP	Multi-Layer Perceptron
NIST	National Institute of Standards and Technology
PCA	Principal Component Analysis
PMC	Perceptron Multi-Couche
POEM	Patterns of Oriented Edge Magnitudes
RAR	Right Acceptance Rate
RBF	Radial Basic Functions
ROC	Receiver Operating Characteristic
RVB	Rouge Vert Bleu
SVM	Support Vectors Machines
SVRG	Speech, Vision and Robotics Group

Chapitre 1

Introduction générale

1.1. Contexte et motivations de la thèse

Le progrès technologique, notamment dans les nouvelles techniques de traitement de l'information et de la révolution imminente des systèmes à microprocesseur ont permis de développer des solutions artificielles pour la résolution et la démystification d'un grand nombre de problèmes causés par l'insécurité et les accès non autorisés dans divers secteurs : détournement de fond, faux et usage de faux, altération ou perte d'informations confidentielles, délits informatiques, ... etc.

L'authentification et l'identification sont des solutions auxquelles on fait recours pour diminuer (voire annuler) les risques de telles situations délictueuses. En effet, l'authentification est la procédure qui a pour but de vérifier et de valider l'authenticité d'une personne ou d'une machine afin que celle-ci puisse accéder à des endroits contrôlés ou à un système d'information. Tandis que l'identification consiste à déterminer de manière à la fois efficace et exacte l'identité d'un individu.

Actuellement, l'authentification des individus est largement généralisée dans tous les domaines nécessitant un accès contrôlé ou sécurisé tels que les applications bancaires, l'accès à des endroits hautement sécurisés comme les sièges de gouvernement, connexion à un ordinateur ou à un réseau informatique, commerce électronique... etc. En revanche, l'identification est souvent utilisée par la police scientifique pour connaître un criminel potentiel et les services d'immigration dans les aéroports... etc.

Souvent pour identifier un individu, on utilise trois méthodes. La première est basée sur une connaissance a priori "**knowledge-based**" de la personne telle que la connaissance d'un mot de passe. La seconde est basée sur la possession d'un objet "**token-based**" telles que la possession d'une pièce d'identité, d'une clef, d'un badge, ...etc. Et la troisième, c'est la combinaison des deux précédentes telle que la carte du guichet automatique bancaire. Ces techniques classées comme des solutions traditionnelles ont d'énormes limites et ont connus en conséquence de lourdes failles. En effet, l'utilisation d'un mot de passe nécessite sa mémorisation, sinon l'écrire ou l'enregistrer qui sont des opérations à hauts risques : la perte ou le vol du code accentue le risque d'être réutilisé par une tierce personne. De même, l'utilisation d'une pièce est elle aussi potentiellement risquée dans la mesure où elle peut être volée et par conséquent falsifiée avec une dextérité et habileté remarquables. C'est dans ce sens même, sens d'une sécurité maximale que les recherches sont orientées vers des méthodes plus sûres, plus fiables et imperméables à tout acte frauduleux de falsification. Grâce à cette orientation, des nouvelles techniques de sécurité ont fait leur apparition et ne cessent de

croître : il s'agit des systèmes biométriques. L'on parle actuellement de la biométrie, une discipline qui recouvre l'ensemble des procédés tendant à établir l'identité ou à vérifier l'authenticité d'un individu à partir de l'analyse des mesures faites sur l'une ou plusieurs de ses caractéristiques physiques, physiologiques ou comportementales.

Depuis des dizaines d'années, la biométrie s'impose comme une technologie privilégiée d'identification des personnes et se trouve actuellement en pleine évolution. En effet, elle fait l'objet d'un essor considérable dans la sphère domestique, dans les établissements scolaires, dans les entreprises, dans le champ de la sécurité, etc. Au départ, elle a été employée au début du siècle passé pour identifier les criminels en s'appuyant sur des mesures anthropométriques du corps caractérisant les personnes. Ensuite, grâce à la découverte de l'unicité de la texture de l'empreinte des doigts pour chaque personne, le champ d'application de la biométrie a été étendu vers l'identification des individus par leurs empreintes auprès des services de police. Mais, la réelle révolution de la biométrie n'a vu le jour qu'après l'apparition d'autres modalités biométriques qui exploitent des caractéristiques physiques de l'homme autres que les empreintes. Ces caractéristiques peuvent être regroupées en trois grandes catégories :

- Les caractéristiques morphologiques telles que : les empreintes digitales, le visage, la géométrie de la main, l'iris, la rétine, l'empreinte palmaire, la géométrie de l'oreille et la voix.
- Les caractéristiques comportementales telles que : la dynamique du tracé de signature, la dynamique de frappe au clavier et la marche.
- Les caractéristiques biologiques telles que : l'ADN, le sang, et la salive.

Les modalités biométriques les plus fréquemment utilisées sont représentées dans la figure 1.1. Certes ces techniques demeurent variables, les unes plus fiables que d'autres, mais il n'en demeure pas moins qu'elles restent infalsifiables et inimitables dans la mesure où chaque individu possède son patrimoine génétique qui lui est propre. Le plus grand avantage, voire le point le plus fort d'exploitation des caractéristiques physiques et physiologiques est leur universalité: mesurables pour tous les êtres humains, uniques permettant de distinguer les personnes entre elles et en même temps permanent (quasi-évolutives).

Pour avoir un regard élargi sur la biométrie: ses racines historiques, ses principaux acteurs, son économie, l'évaluation des dispositifs biométriques et ses usages, le lecteur peut se référer aux ouvrages [1] [2] [3], les publications [4] [5] et le site web¹.

1. www.biometrie-online.net




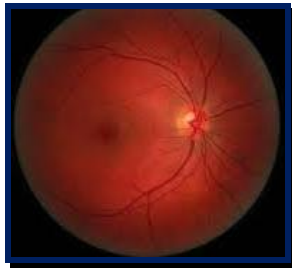





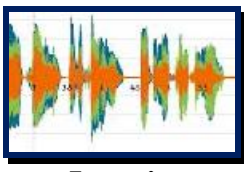
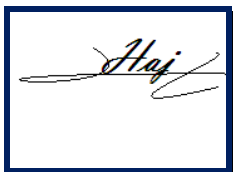

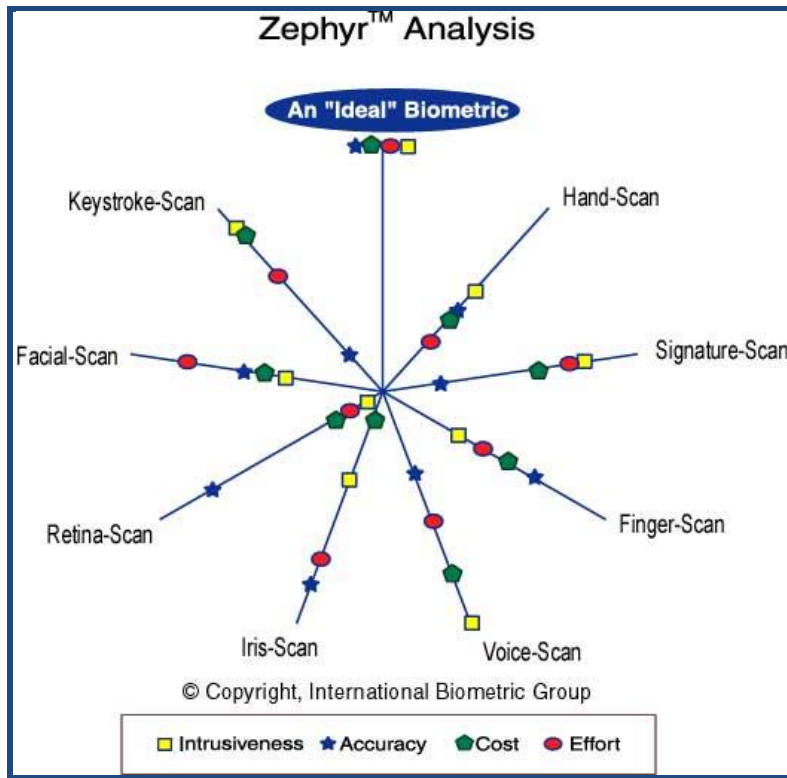
			
L'empreinte digitale	La géométrie de la main	L'iris	La rétine
			
La démarche	L'empreinte palmaire	La géométrie de l'oreille	L'ADN
			
Le visage	La voix	La signature	La dynamique de frappe

Fig. 1.1. Principales modalités biométriques.

Le dénominateur commun des technologies biométriques est d'offrir des moyens simples, fiables, pratiques et moins coûteux pour vérifier l'identité d'une personne sans l'assistance d'une autre personne. Mais, jusqu'à présent, aucune modalité biométrique n'a pu garantir la totalité de cet objectif car aucune d'elles n'est optimale. Alors, devant sa diversité, l'analyse comparative de leurs performances est importante avant la prise de décision sur la modalité à implanter car le choix dépend des fonctions recherchées par l'application visée.

Plusieurs études ont été menées afin de comparer les performances des systèmes biométriques pour orienter les entreprises spécialisées dans la biométrie lors du choix d'une technologie en fonction des contraintes liées à l'application. Par exemple, la société américaine IBG² (*International Biometric Group*) a procédé à une comparaison de huit technologies sur la base de quatre critères d'évaluation : L'effort, l'intrusivité, le coût et la fiabilité. Cette comparaison est publiée sous forme d'une analyse Zéphyr (figure 1.2).

2. www.biometricgroup.com/



<i>Légende</i>
Effort : effort requis pour l'utilisateur lors de la mesure.
Intrusiveness : décrit dans quelle mesure l'utilisateur perçoit le test comme intrusif.
Cost : coût de la technologie (lecteurs, capteurs, etc...)
Accuracy : efficacité de la méthode (capacité à identifier quelqu'un)

Fig. 1.2. Analyse Zephyr (source: IBG)

La reconnaissance de visages qui est classée parmi les techniques principales de la biométrie est facilement tolérée par les utilisateurs, peu intrusive et moins coûteuse, à condition que des taux de reconnaissance suffisants soient garantis. Cette spécificité de la biométrie de visage est confirmée par l'étude comparative d'IBG ci-dessus. Après l'analyse des résultats de cette étude, nous pouvons conclure que la reconnaissance de visage est:

- la plus facile à utiliser car elle est non contraignante. En effet, elle n'exige pas beaucoup d'effort de la part de l'utilisateur lors de la saisie de mesures. Certains systèmes employant cette technique fonctionnent même sans que l'utilisateur en soit conscient.
- une technique non intrusive car on ne compte pas sur une coopération active de la personne à reconnaître. En effet, la capture d'image de visage est sans contact à l'instar d'autres modalités telle que la rétine qui est difficile à capter et exige un contact direct entre le capteur utilisé et l'individu à reconnaître. Ce qui rend la modalité de visage bien acceptée par les utilisateurs.

De plus cette modalité a :

- un coût modéré car elle ne demande pas d'équipement d'acquisition très complexe et coûteux. En effet, la plupart des systèmes fonctionnent en utilisant une camera numérique de qualité normale. De plus, Cette approche recevant une information de nature optique et les

techniques de traitement optique de l'information auront dans le futur proche un grand intérêt.

- une fiabilité controversée et son taux de reconnaissance n'est pas très élevé car cette modalité reste encore relativement sensible à l'environnement réel d'acquisition d'image tels que la variation d'illumination et l'angle de prise de vue qui peuvent dégrader la performance de fiabilité.

La reconnaissance de visage avec ses forces et ses faiblesses se trouve actuellement en perpétuel développement. Elle a suscité une attention substantielle dans les dernières décennies de la part des groupements de recherche académiques et les entreprises spécialisées dans la biométrie et la sécurité. Cette attention se justifie pour la première communauté, par le nombre important de travaux publiés sous le thème de la reconnaissance de visage car elle est considérée comme un axe de recherche très attractif. Tandis que, pour la deuxième communauté, par le grand nombre de produits commerciaux de la reconnaissance de visage vendus dans le marché. Ces produits connaissent une utilisation à grande échelle en deuxième position après les empreintes. Cet emplacement est affirmé par les statistiques faites sur le marché de l'industrie biométrique publiées par IBG (*International Biometric Group*) dans son rapport (2009-2014). Ce rapport confirme que la technologie des empreintes digitales devrait gagner **28,4%** du marché de la biométrie, suivie de la reconnaissance de visage à **11,4%**. L'histogramme de ces statistiques est représenté dans la figure 1.3.

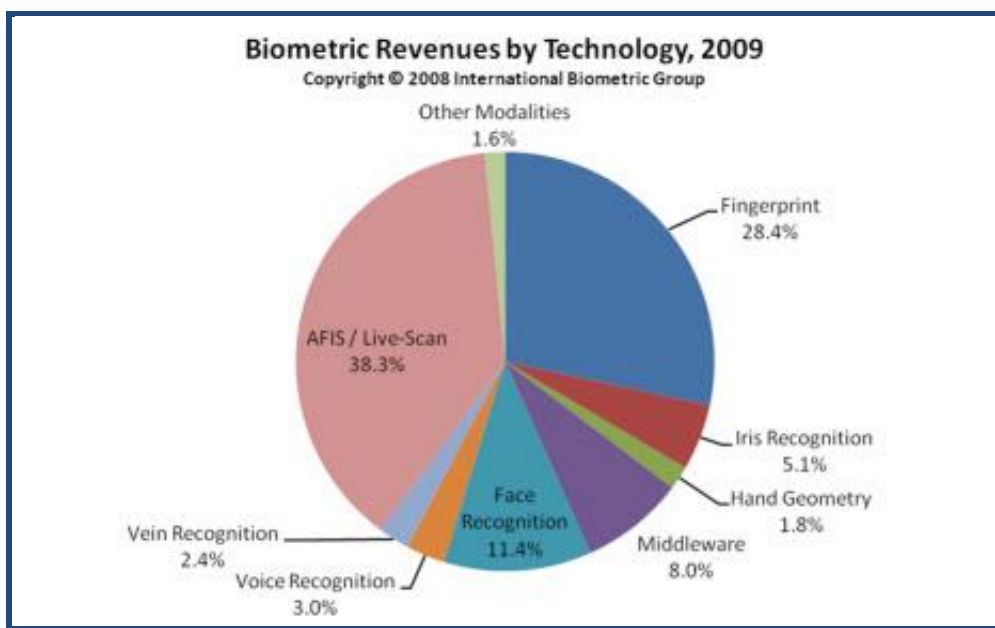


Fig. 1.3. Statistiques des revenus des technologies biométriques 2009-2014

En plus des deux communautés précédentes (chercheurs et entreprises), s'ajoute un troisième acteur qui a énormément contribué au développement de cette technologie, ce sont les organisations gouvernementales et industrielles. Ces dernières ont lancé des programmes et des projets dans le but de mesurer les progrès réalisés dans ce domaine et d'identifier les orientations futures de la recherche pour que la reconnaissance faciale soit mieux déployée.

Le premier programme qui s'intéressait aux tests des algorithmes de la reconnaissance de visage, c'était FERET (*Face Recognition Technologie*) [6] [7] sponsorisé à partir de Septembre 1993 par le ministère de la défense américaine DOD et exactement par l'agence DARPA (*Defense Advanced Research Projects Agency*) et supervisé par le laboratoire américain ARL (*Army Research Laboratory*). Ce programme était composé de trois éléments majeurs, le premier était la sponsorship des laboratoires de recherche qui étudient la théorie de la reconnaissance de visage, le deuxième était la collecte et la distribution d'une base de données standard des images de visage références connue ensuite par le nom de FERET, et le dernier élément était les évaluations FERET qui ont testé les performances des algorithmes de reconnaissance faciale les plus populaires à l'aide de cette base. Cette dernière est actuellement maintenue à l'institut NIST³ (*National Institute of Standards and Technology*).

Après le programme FERET, les mêmes organismes (DARPA et NIST) ont lancé le projet FRVT (*Face Recognition Vendor Test*) à partir de 2000. FRVT a conçu des évaluations indépendantes des technologies de reconnaissance faciale et prototypes disponibles dans le marché. Les statistiques de ces évaluations ont été publiées ensuite sous forme de rapports d'évaluation en commençant par le premier rapport de FRVT 2000 [8] jusqu'au dernier rapport de FRVT 2013 [9]. En sus des résultats des évaluations, les rapports ont évoqué les difficultés de la reconnaissance de visage et ils ont conclu en mettant en relief un ensemble de recommandations.

Autres projets dont les objectifs sont proches à ceux de FERET et FRVT sont annoncés pour promouvoir la technologie de visage fusionné avec une autre modalité biométrique, tels que : MBGC (*Multiple Biometric Grand Challenge*), FOCS (*Face and oculaire Challenge Series*) et MBE (*Évaluation des données biométriques multiples*).

1.2. Objectifs de la thèse

La mise en œuvre d'un système de détection et de reconnaissance de visages répondant à des critères de fiabilité, de robustesse (taux de reconnaissance très élevé dans n'importe quel environnement) et de rapidité (temps de reconnaissance très faible pour assurer une

3. <http://www.nist.gov/>

reconnaissance en temps réel), reste un problème très ouvert de nos jours. En effet, à l'heure actuelle, il existe de nombreux algorithmes de reconnaissance de visage performants sous réserve que les conditions d'acquisition des visages soient contraintes, ce qui nécessite que les sujets soient coopératifs et veuillent se faire identifier. Mais les problèmes de reconnaissance de visages en environnements non contrôlés (le cas de la vidéosurveillance) ne sont pas encore résolus.

Notre thèse s'inscrit dans une telle problématique. Elle a pour objectif la conception d'une architecture logicielle complète d'un système automatisé, robuste et rapide destiné à la détection et la reconnaissance de visages. Afin d'atteindre ce but, nous avons travaillé à tous les niveaux de la chaîne de traitement d'un tel système et nous avons apporté des contributions pertinentes à ses différents points pivots, notamment: la détection et la localisation de visage, l'extraction des vecteurs caractéristiques de visage (signature de visage), l'apprentissage et la classification de ces vecteurs.

1.3. Principales contributions

La première contribution de cette thèse a consisté dans le développement d'une nouvelle approche pour la détection et la localisation de visage dans une image fixe à multiple visages. Cette approche est de nature hybride puisqu'elle utilise dans son traitement en premier temps une caractéristique invariante de visage, la couleur de peau comme information de segmentation d'une image couleur, et en deuxième temps un modèle (Template-matching) d'un visage type pour la classification visage et non-visage des régions de peau. Dans la phase de segmentation, nous avons introduit un nouveau modèle de la représentation de la distribution de la couleur de peau pour détecter les pixels de peau dans une image couleur. Puis, nous avons procédé à une segmentation en régions de peau connexes à l'aide d'un algorithme très efficace en segmentation, la Ligne de Partage des Eaux (LPE) après son amélioration. Cette dernière avait l'objectif d'approcher sa définition informatique à celle naturelle d'origine. Tandis que, pour le modèle de visage élaboré, nous l'avons défini sous forme d'un ensemble de règles à partir de l'apparence globale et locale d'un visage type. Ces règles qui ont montré leur capacité de classer une image-candidate de visage en deux classes (visage et non-visage), sont appliquées sur la réponse de la convolution de l'image région de peau avec une cascade de filtres de Gabor. Les résultats obtenus avec cette approche prouvent son efficacité en termes de fiabilité et son caractère de rapidité.

La deuxième contribution de ce travail est le développement d'une nouvelle méthode pour extraire des images de visage les vecteurs caractéristiques (signatures) ACP-ADL double :

l'ACP (Analyse en Composantes Principales, "*EigenFaces*") suivie de l'ADL (Analyse Discriminante Linéaire) doublée deux fois successives. L'objectif visé par la combinaison de ces deux algorithmes, c'est de préserver les avantages des deux. L'ACP permet de réduire la dimensionnalité de l'espace de représentation de l'image de visage tout en optimisant sa reconstruction. Tandis que, l'ADL a une puissance de discrimination entre les classes de visage après projection. Et pour garantir de plus cette puissance, nous avons doublé la projection avec l'ADL. Cette méthode abouti à des performances très satisfaisantes en termes de rapidité dans la phase de reconnaissance et du taux de reconnaissance lorsqu'elle est testée sur des images de visages qui ne possèdent pas de fortes variations d'expression et d'illumination et de pose. D'autre part, cette méthode tolère les occultations partielles et la variation d'échelle.

La troisième contribution de cette thèse est élaborée dans le sens de surmonter les problèmes de la première approche d'extraction des caractéristiques (signatures), à savoir : la chute du taux de reconnaissance lorsque nous présentons au système des images de visage ayant des fortes variations de pose et d'illumination. Cette contribution consiste dans le développement d'une nouvelle approche d'extraction des caractéristiques de nature hybride. Dans cette approche, nous avons introduit en premier lieu le descripteur de Gabor pour la modélisation de l'image de visage. Ce type descripteur assure une minimisation des effets liés aux variations d'illumination et de pose. En deuxième lieu, nous avons intégré un estimateur de pose entre la phase d'extraction et celle de classification pour faciliter la reconnaissance de visages pris dans des poses variables.

La quatrième contribution est d'avoir présenté une synthèse des méthodes de classification classiques et modernes les plus utilisées en reconnaissance de visages : les distances métriques, les réseaux de neurones et les Séparateurs à Vastes Marges (SVM)

1.4. Organisation de la thèse

Ce manuscrit doit pouvoir montrer clairement le cheminement scientifique qui s'est effectué tout au long de cette thèse, depuis l'imprégnation de l'état de l'art jusqu'à l'évaluation des approches proposées pour la mise au point de notre système de reconnaissance de visages. Pour atteindre cet objectif, nous avons choisis d'articuler notre manuscrit autour de huit chapitres organisés de la manière suivante :

- **Chapitre 1:** c'est le présent chapitre dans lequel nous avons introduit les grandes lignes du contexte de notre thèse, à savoir : la problématique de la sécurité et sa résolution au moyen des technologies biométriques, en particulier, la reconnaissance de visages. Ensuite, nous

avons présenté la position de cette modalité de visage par rapport aux autres techniques biométriques principales et les facteurs qui nous ont motivé pour choisir cet axe de recherche. Enfin, nous avons présenté les objectifs de notre thèse.

- **Chapitre 2:** le but de ce chapitre est de donner un panorama de la reconnaissance de visages, en soulignant en premier temps ses intérêts et ses champs d'application. Ensuite, une description détaillée de l'architecture de base d'un système automatisé de reconnaissance de visages qui dévoile ses différents modules. A travers ce panorama, nous voulons également mettre en lumière les challenges de la reconnaissance de visages, puis les outils mis en œuvre pour évaluer ses performances.
- **Chapitre 3:** dans ce chapitre de l'état de l'art en reconnaissance de visages, nous évoquerons les grandes lignes de l'évolution de la reconnaissance des visages. Tout d'abord, une brève présentation du principe des méthodes qui exploitent le modèle 3D de visage. Après, une étude et une analyse des méthodes les plus populaires qui ont marqué le domaine de la reconnaissance faciale 2D. Cette analyse vise à classifier ces techniques de manière efficace et à mettre systématiquement l'accent sur les particularités, les avantages et les inconvénients de chacune d'elles. Ceci nous aidera dans le choix du sens de recherche à suivre.
- **Chapitre 4:** dans ce chapitre, nous nous intéressons à la présentation détaillée de notre nouvelle approche pour la détection et la localisation de visage dans une image fixe à multiple visages. cette présentation est précédée par un survol de la classification des méthodes utilisées dans la détection de visage.
- **Chapitre 5:** ce chapitre est consacré à la phase d'extraction des caractéristiques (vecteurs caractéristiques). Les deux nouvelles approches mises en œuvre pour cette tâche seront présentées avec une étude approfondie de tous les outils exploités pour les développer.
- **Chapitre 6:** dans ce chapitre, nous nous concentrons sur les méthodes de classification. Nous étudierons en détail les méthodes les plus utilisées en reconnaissance de visage, à savoir : les distances métriques, les réseaux de neurones et les Supports Vecteurs Machines ou Séparateurs à Vaste Marge (SVM). Cette étude sera suivie par une synthèse expérimentale dans le chapitre suivant.
- **Chapitre 7 :** ce chapitre fera l'objet de la présentation des résultats expérimentaux obtenus après les tests de nos approches proposées sur des bases de données standards. Nous discutons les performances de ces approches et tirons une conclusion sur l'utilisation du système de reconnaissance de visage tout entier dans un environnement non-contrôlé.

- **Bilan et perspectives :** Ce dernier chapitre conclut ce manuscrit en résumant nos principaux apports et présente les perspectives ouvertes par les travaux de cette thèse.

Chapitre 2

La reconnaissance de visages

2.1. Introduction

Pour développer les systèmes d'intelligence artificielle, les scientifiques s'inspirent généralement du comportement naturel de l'être humain, et c'est le cas de la reconnaissance de visages qui constitue une des capacités les plus étonnantes du système visuel humain.

Le visage qui fait partie des caractéristiques biométriques est l'un des objets visuels les plus importants de notre environnement, auquel nous sommes exposés depuis notre naissance. Certains auteurs [10] nous confirment que les particularités morphologiques observées sur les différentes régions faciales, comme : les yeux, le nez, la bouche et les sourcils, montrent une bonne capacité discriminante lorsqu'elles sont utilisées par les différentes analyses anthropologiques. Cette particularité du visage a développé chez l'être humain l'aptitude de discriminer des milliers de visages, d'en mémoriser des centaines, d'en identifier beaucoup. Le but ultime de la reconnaissance artificielle de visages est de rivaliser, voire dépasser ces capacités humaines de reconnaissance.

Dans ce chapitre, nous allons décrire en détail l'architecture de base de tout système artificiel de reconnaissance de visages. Mais, avant cette description, nous allons lister les champs d'application de cette modalité biométrique et ses modes de fonctionnement. Ensuite, nous allons présenter ses challenges majeurs et les outils mis en œuvre pour évaluer un tel système.

2.2. Applications de la reconnaissance de visages

La reconnaissance de visages est un axe de recherche ouvert attirant des chercheurs relevant de disciplines différentes : la psychologie, le traitement d'image, l'intelligence artificielle, la robotique, la vision par ordinateur, l'infographie et les réseaux informatiques ... etc. Cette diversité dans les disciplines de recherche a fait éclore un grand nombre d'algorithmes de reconnaissance de visages. Et grâce à ces derniers et à la révolution imminente des systèmes à microprocesseur, l'industrie a mis au point de nombreux produits commerciaux destinés aux domaines suivants:

- La sécurité : apporter un niveau de sécurité supérieur en utilisant le mode d'identification dans les lieux qui nécessitent un contrôle d'accès physique (bâtiments sécurisés, aéroports, ports maritimes, points de contrôle des frontières, etc.) et dans les systèmes d'information qui demandent un contrôle d'accès logique (ordinateurs, réseaux informatiques, bases de données sensibles, e-commerce, guichets automatiques, téléphones mobiles, etc....).

- La surveillance : la recherche dans des endroits publics des criminels, des terroristes connus, des délinquants toxicomanes, des enfants disparus, des immigrants pour que les autorités puissent les notifier là où elles les retrouvent.
- La vérification de l'identité : telle que la carte d'identité nationale, le permis de conduite, la sécurité sociale, le passeport, etc.
- A part ces applications qui utilisent le visage pour l'identification ou la vérification d'identité, il y a d'autres alternatives de cette modalité, telles que : l'indexation de la vidéo par visage, le filtrage des sites Web par critère de visage, la reconstruction du visage témoin, les applications *Facial Network* qui permettent d'assembler tous les profils d'une personne sur les réseaux sociaux disponibles sur internet, etc.

2.3. Modes de fonctionnement de la reconnaissance de visages

Lors de l'implantation d'un système automatisé de reconnaissance de visages, le contexte de son champ d'application fixe son mode de fonctionnement à adopter parmi les deux modes suivants :

- Le mode de vérification ou authentification (un à un) : le système confirme l'identité d'une personne proclamée en comparant la signature de l'image capturée de son visage avec celui de cette même personne stockée dans la base de données du système. Dans un tel mode, le système doit alors vérifier si la personne est bien celle qu'elle prétend être ou non ; accepté en tant qu'utilisateur légitime ou refusé en tant qu'imposteur.
- Le mode d'identification (un à N) : le système doit affecter une identité de la personne qui se présente devant lui sans aucune connaissance a priori sur son identité. Dans ce cas, le système compare la signature de l'image capturée de son visage avec toutes les N signatures des personnes connues enregistrées dans la base de données (galerie) du système pour lui associer « une identité » ou l'étiquette « inconnu ».

2.4. Architecture de base d'un système de reconnaissance de visages

L'être humain est capable de discriminer un très grand nombre de visages, même s'ils partagent un ensemble de traits similaires. Cette perception visuelle est un processus cognitif qui repose sur des mécanismes neurocognitifs complexes par lesquels le cerveau analyse une scène pour y détecter et identifier un visage [11]. Mais, pour que notre système visuel soit capable d'identifier une personne à partir de son visage, il a besoin d'avoir préalablement associé et mémorisé l'allure de son visage avec son identité. Cette opération appelée apprentissage, est réalisée facilement chez les êtres humains d'une manière spontanée et

évolutive. Elle doit être intégrée dans un système artificiel de reconnaissance de visages et effectuée d'une manière supervisée.

Le processus d'apprentissage ou enrôlement constitue le premier module qui doit être exécuté dans un système de reconnaissance de visages. Tandis que le deuxième module est celui de la reconnaissance (vérification ou identification). Le premier module sert à extraire pour chaque personne souhaitant utiliser le système ultérieurement, un vecteur caractéristique (signature) à partir de l'image de son visage capturée. L'ensemble des vecteurs caractéristiques extraits des personnes connues par le système sont ensuite stockées dans une base de données de références (galerie). Alors qu'en module de reconnaissance, après capture de l'image ou la vidéo contenant le visage de la personne à reconnaître, il faut extraire son vecteur caractéristique (signature) pour aborder la phase de comparaison (classification) qui permet de fournir la décision finale sur son identité.

Le module d'enrôlement a le même principe quelque soit le type du mode de fonctionnement de l'application (vérification ou identification) et son traitement n'a pas de contrainte de temps, puisqu'il s'effectue « hors-ligne ». Par contre, dans le module de reconnaissance, la méthode de classification dépend de l'application envisagée (vérification ou identification) et son traitement s'effectue en « ligne ». Les étapes de traitement à parcourir séquentiellement dans ces deux modules sont illustrées dans la figure 2.1 et décrites ci-dessous. Les étapes 1 à 5 doivent être exécutées dans les deux modules.

Etape 1 : L'acquisition d'image (vidéo)

C'est la première étape dans le processus des deux modules qui permet de capturer la scène contenant un visage à l'aide d'un capteur d'image (vidéo) : caméra, appareil photo,...etc. On fait l'acquisition d'image (image fixe ou vidéo) du monde physique pour la transférer vers l'unité de traitement. Elle est une étape très importante dans les systèmes de reconnaissance de visages. C'est pour cette raison qu'il faut réussir à capter des images sans bruit et de bonne qualité pour améliorer les performances de reconnaissance. Il existe plusieurs types de capteurs pour l'acquisition du visage qui se classe selon leur mode de fonctionnement, leur domaine de sensibilité spectrale et leur mode d'acquisition.

Dans le domaine du visible, il existe aujourd'hui quatre grandes technologies de capteurs : CCD (*Charge Coupled Devices*), CMOS (*Complementarity Metal Oxide Semiconductor*), Super CCD et Foveon. Les plus répandus actuellement, tant sur les boîtiers professionnels qu'amateurs en petit format sont les CMOS. Par ailleurs, il existe actuellement des systèmes qui utilisent des capteurs d'images thermiques ou infrarouges.

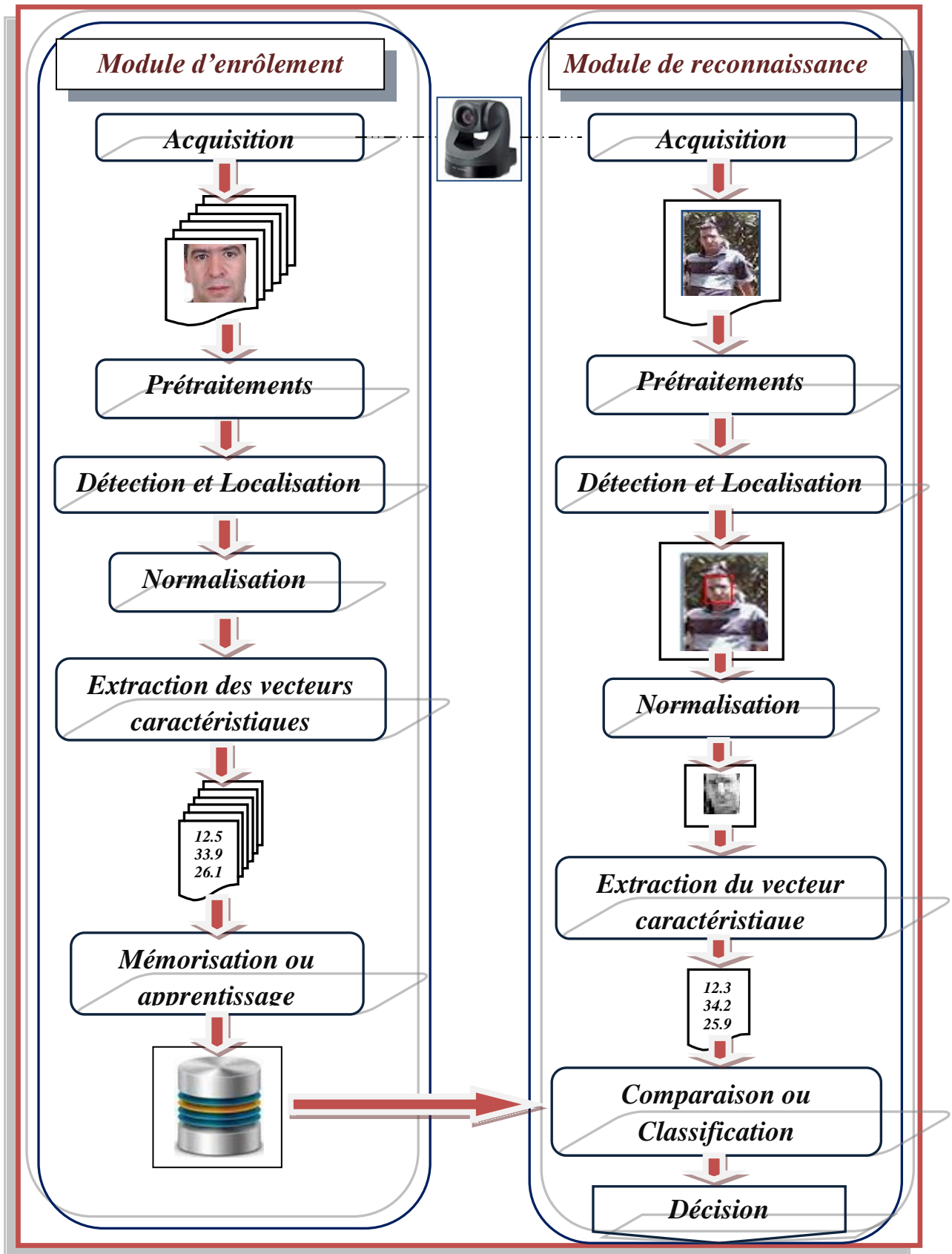


Fig. 2.1. Architecture d'un système de reconnaissance de visages

Etape 2: Le prétraitement

Après avoir capturé l'image (vidéo), il faut consulter sa qualité pour déterminer son aptitude pour le traitement ultérieur. Si cette qualité est insuffisante, il faut l'améliorer par des prétraitements pour augmenter sa compacité et diminuer le bruit. L'élimination du bruit se fait avec des techniques de traitement d'images et le rétablissement de la qualité initiale de l'image est réalisé par des techniques de restauration d'images.

Etape 3 : La détection et la localisation

L'image (vidéo) acquise peut contenir à la fois le visage de la personne à reconnaître et éventuellement un arrière-plan. Dans cette étape, on détecte la présence ou l'absence du visage dans l'image acquise. Si l'image contient un visage, on localise son emplacement pour l'extraire. Cette étape est autant plus délicate autant que l'image acquise contient plusieurs objets de visage. C'est une étape cruciale dans un système de reconnaissance de visages et la qualité de l'algorithme de son traitement a un impact direct sur la fiabilité et la robustesse du système.

Etape 4 : Normalisation

La normalisation a pour objectif d'éliminer ou réduire les influences des variations d'échelle, de pose et d'illumination dues aux changements des conditions d'acquisition de l'image. Pour les deux premiers types de variation, la normalisation a un aspect géométrique car elle ajuste les dimensions et les positions des visages localisés pour qu'ils aient une taille commune et qu'ils soient horizontales. Alors que pour le troisième type de variation, l'image subit une normalisation photométrique pour en uniformiser l'éclairage par des méthodes simples telle que l'égalisation d'histogramme, la correction gamma ou par des méthodes plus complexes telle que le lissage.

Etape 5 : L'extraction des caractéristiques

C'est l'étape clé du processus, car la performance du système entier en dépend. Dans cette étape appelée aussi indexation ou modélisation, on extrait de l'image de visage les informations les plus pertinentes qui permettent de modéliser le visage d'une personne par un vecteur de mesure (vecteur caractéristique ou signature) qui le caractérise.

Le choix judicieux de la méthode d'extraction est extrêmement important dans la mesure où l'étape suivante (classification ou comparaison) se fera uniquement sur la base des valeurs de ces vecteurs. Par conséquent ces derniers doivent posséder les propriétés suivantes:

- La discrimination : ils doivent prendre des valeurs significativement différentes pour des visages de personnes différentes ; Pour une personne donnée, sa signature ne doit pas ressembler à celle d'une autre personne.
- La robustesse à toutes les variations possibles, telles que l'expression, la pose, l'illumination ; Pour une personne donnée, sa signature ne doit pas varier avec le changement des conditions d'acquisition.
- La Taille réduite : pour diminuer la quantité de données à stocker car la complexité d'un système de reconnaissance des visages augmente rapidement avec la taille de ces vecteurs.
- Le temps de leur extraction doit être faible, surtout dans le module de reconnaissance.

Généralement, ce sont les approches déployées dans les étapes de détection de visage et d'extraction des caractéristiques qui font la distinction entre les systèmes de reconnaissance de visages.

Etape 6 : Mémorisation des vecteurs caractéristiques

Dans cette étape, on mémorise les vecteurs caractéristiques extraits des individus connus. Pendant ce processus de mémorisation, chaque vecteur stocké dans la base de données est associé à une certaine identité personnelle telle que le nom, numéro d'identification personnel qui caractérise son utilisateur. Cette base de données référentielle est centralisée sur un serveur central ou décentralisée sur une carte à puce que seule la personne à reconnaître possède en fonction de l'application envisagée.

Etape 7 : Classification ou comparaison

Cette étape s'exécute uniquement dans le module de reconnaissance et son traitement peut être réalisé par deux alternatives. Dans la première, le vecteur caractéristique de la personne à reconnaître subit une comparaison, soit avec celui de référence de la même personne issue de la base de données dans le cas de vérification d'identité ou avec tous les vecteurs références de la base de données dans le cas d'identification. Cette comparaison se fait à l'aide d'un algorithme de calcul de similarité ou un algorithme de calcul de distance. Tandis que, pour la deuxième alternative, le vecteur caractéristique est soumis à l'entrée d'un classificateur utilisant l'une des méthodes modernes de classification des données, telles que : les réseaux de neurones et les machines à vecteurs de support (*Support Vectors Machines, SVM*). Dans ce cas l'étape précédente se matérialise par un apprentissage supervisé du classificateur utilisé.

Étape 8 : Décision

Dans cette étape finale, on doit aboutir à une décision sur l'identité de la personne à reconnaître. Pour le faire dans le mode vérification d'identité, le score de similarité ou de distance produit par l'étape de comparaison, est comparé à un seuil de décision fixé en avance pour répondre à la question: acceptée comme " utilisateur légitime" ou rejetée comme un "imposteur". Pour le mode d'identification, l'identité de la personne de la base de référence qui donne le plus haut score de similitude (ou le petit score de distance) est assignée à la personne à identifier. Dans ce mode, pour résoudre le problème des individus non enregistrés dans la base de données et qui doivent être étiquetés " **non inconnu**", on compare le score avec un seuil. Par ailleurs, si on utilise un classificateur (réseaux de neurones ou machines à vecteurs de support), c'est sa sortie qui indique à quelle classe appartient le visage à reconnaître.

2.5. Challenges liés à la reconnaissance de visages

Malgré les progrès très conséquents dans le domaine de la reconnaissance de visages, des algorithmes génériques de reconnaissance n'ont toujours pas été dégagés. En effet, il reste de nombreux challenges à surmonter. Notamment, dans un environnement non contrôlé où l'apparence des visages à reconnaître varie énormément, et ceci peut aussi bien être lié aux conditions dans lesquelles sont acquises les images qu'à la qualité de l'acquisition elle-même. Cette variabilité n'est pas tolérée par la majorité des solutions proposées qui fonctionnent uniquement dans des environnements contrôlés. C'est à dire que toutes les conditions pouvant dégrader les performances attendues par ses solutions sont maîtrisées. Les individus sont coopératifs si bien que les visages sont acquis de face, sans expression ni variation d'illumination. Les évaluations élaborées dans les programmes FERET et FRVT et les résumés de l'état de l'art [12] [13] [14] ont révélé que les problèmes de variation d'illumination et de pose constituent un défi majeur pour la reconnaissance de visage. En général, la variabilité de l'apparence d'un visage est liée à des facteurs multiples difficilement modélisables. Ces facteurs peuvent avoir un aspect extrinsèque ou intrinsèque.

2.5.1. Facteurs extrinsèques

2.5.1.1. Variation d'illumination

La variation des caractéristiques de la source de lumière (son intensité et son spectre), la distribution de l'éclairage, la lumière d'angle et les caractéristiques de l'appareil d'acquisition d'image changent radicalement l'apparence d'un visage (figure 2.2). Elles

peuvent entraîner une mauvaise classification des images d'entrée issues de la même classe (la même personne). Deux types d'éclairage peuvent influencer l'apparence d'un visage: l'illumination globale (ou ambiante) qui affecte tout le visage de manière uniforme (ou presque) et l'illumination locale qui entraîne la création d'ombre et des zones éclairées de manière non linéaire. De nombreux algorithmes ont été proposés pour palier les problèmes d'illumination [15] [16].

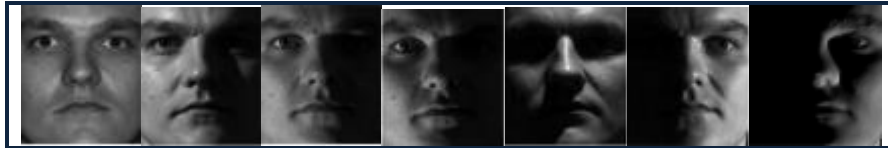


Fig. 2.2. Exemples de visages avec variation d'illumination

2.5.1.2. Variation de pose

La pose d'un visage définit la rotation qu'à pu subir un visage lors de la capture. Les variations de pose peuvent être de deux types selon l'axe de rotation : rotation dans le plan où l'axe de rotation est l'axe de la caméra, et rotation hors plan sinon. La figure 2-3 présente des exemples d'un visage subissant une variation de pose. Cette dernière est une difficulté majeure qui peut baisser considérablement le taux de reconnaissance de visage. En effet, Quand le visage est de profil dans le plan image (orientation $< 30^\circ$), tous les traits caractéristiques du visage (les yeux, le nez et la bouche) peuvent être détectés. Cependant, lorsque la rotation est supérieure à 30° , la détection totale n'est plus possible. Vu ce problème, un grand nombre de systèmes de reconnaissance automatique de visages se limitent aux poses frontales, ou à des poses spécifiques. Dans le cas d'une rotation dans le plan, l'apparence du visage n'est pas déformée, et une bonne estimation de l'angle de la rotation peut suffire à recalculer l'image par simple rotation inverse, et ainsi obtenir une pose frontale. Le cas de la rotation hors plan est souvent bien plus complexe, sauf si les visages utilisés pour l'enrôlement et la reconnaissance présentent la même pose. Dans les résumés [14] [17] on trouve une synthèse qui présente les algorithmes de reconnaissance de visages les plus connus qui sont sensibles à la variation de pose et les récentes approches développées pour réduire l'effet de cette variation. Les problèmes liés à l'orientation du visage par rapport à la caméra restent le défi majeur des différentes techniques de reconnaissance actuelles, c'est pour cette raison qu'un nombre important de chercheurs sont orientés vers la résolution de ce problème [18] [19] [20].



Fig. 2.3. Exemples de visages avec variation de pose

2.5.1.3. Occultations partielles

Les visages peuvent être partiellement occultés par d'autres objets dans la scène. Ces occultations partielles apparaissent fréquemment dans des applications réelles, comme illustré sur la figure 2.4. Elles peuvent être causées par une main cachant une partie du visage, par des cheveux longs, des lunettes de vue, de soleil, par tout autre objet (foulard . . .), ou encore par d'autres visages dans le cas d'une image contenant un groupe de personnes. Parmi les approches proposées dans le sens de palier ce problème, on trouve : [21] [22] [23]. Par ailleurs, une présentation d'autres méthodes de ce genre avec leur comparaison est détaillée dans la récente synthèse élaborée dans [24].



Fig. 2.4. Exemples de visages avec des occultations partielles

2.5.1.4. Bruit introduit par le capteur

Le bruit peut être présent dans l'image de visage acquise, ceci étant principalement dû à un capteur défaillant ou mal entretenu. Par exemple, un mauvais focus de caméra entraînant du flou dans les images captées.

2.5.2. Facteurs intrinsèques

2.5.2.1. Expressions faciales

Un autre facteur qui affecte l'apparence du visage est l'expression faciale (figure 2.5). Les traits caractéristiques tels que la bouche, les yeux ou encore les sourcils peuvent alors subir des déformations importantes, pouvant faire échouer un système de reconnaissance de visages fondée sur la géométrie de ces points caractéristiques. La reconnaissance de visage avec une variation de l'expression faciale est un problème à résoudre car il entraîne forcément une

diminution du taux de reconnaissance. On trouve dans la littérature des travaux qui ont développé des algorithmes spécifiques à ce genre de variation, tels que [23].

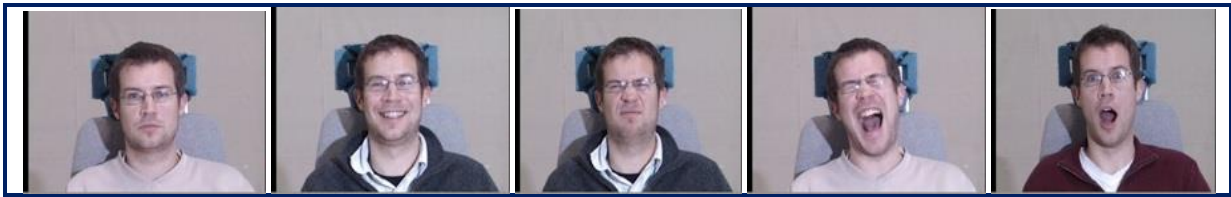


Fig. 2.5. Exemples de visages avec variation de l'expression faciale

2.5.2.2. Présence ou absence des composants structurels

La présence des composants structurels telle que la barbe ou la moustache peut modifier énormément les caractéristiques faciales telles que la forme, la couleur, ou la taille du visage. De plus, ces composants peuvent cacher les caractéristiques faciales de base causant ainsi une défaillance du système de reconnaissance.

2.6. Outils d'évaluation d'un système de reconnaissance de visages

Dans le chapitre 1, pour situer la technologie de visage par rapport aux autres modalités biométriques, nous avons exploité l'étude comparative réalisée par IGB en se basant sur les quatre critères (effort, intrusivité, coût et fiabilité). Dans cette section, nous présenterons les outils d'évaluation standards les plus utilisés par la communauté pour classer les systèmes de reconnaissance de visage entre eux en termes de fiabilité. Cette performance est évaluée différemment selon que l'on considère un mode de fonctionnement en vérification ou en identification car les taux à mesurer sont spécifiés par le mode adopté.

2.6.1. Le mode de vérification

Pour le mode de vérification, le système doit déterminer si l'identité proclamée par un utilisateur est acceptée (un utilisateur légitime) ou rejetée (un imposteur). Dans les deux cas, lors de la prise de la décision, deux erreurs peuvent surgir: la fausse acceptation d'un imposteur qui doit être rejeté ou le faux rejet d'un utilisateur légitime qui doit être accepté. Dans la figure 2.6, nous présentons la distribution hypothétique des scores de similarité qu'obtiendraient les utilisateurs légitimes et les imposteurs. Cette figure montre bien l'influence de la valeur du seuil de décision sur les performances du système. Un seuil trop petit entraîne l'apparition d'un grand nombre de fausses acceptations, tandis qu'un seuil trop grand engendre un grand nombre de faux rejets.

En partant des deux erreurs signalées précédemment, on calcule deux taux considérés comme un critère d'évaluation de la fiabilité d'un système de vérification d'identité par visage:

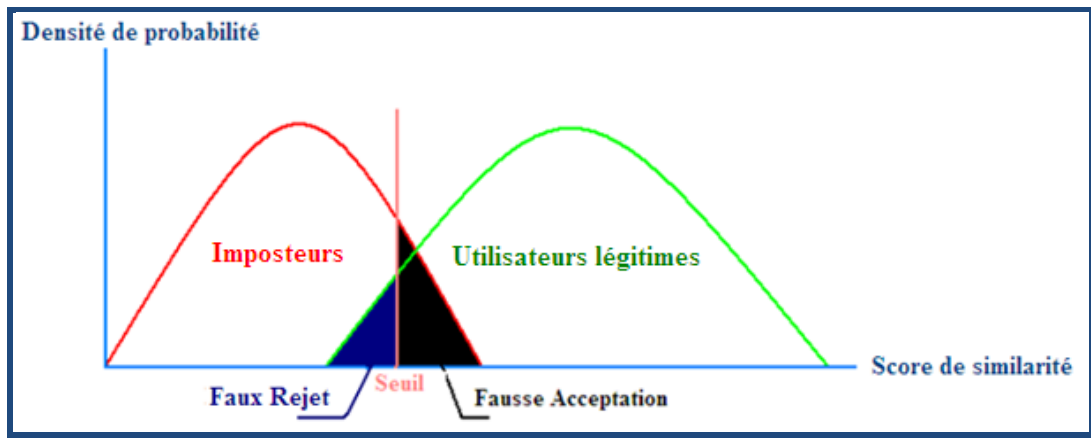


Fig. 2.6. Illustration des erreurs : Faux Rejet et Fausse Acceptation

- Le Taux de Fausses Acceptations FAR (*False Acceptance Rate*): ce taux exprime le pourcentage d'utilisateurs acceptés par le système alors qu'ils devraient être rejetés (des imposteurs) en utilisant la formule suivante :

$$FAR = \frac{\text{Nombre de fausses acceptations}}{\text{Nombre total d'accès imposteurs}} \quad (2.1)$$

- Le Taux de Faux Rejets FRR (*False Rejection Rate*): ce taux exprime le pourcentage d'utilisateurs rejetés alors qu'ils devraient être acceptés (des utilisateurs légitimes) par le système en utilisant la formule suivante :

$$FRR = \frac{\text{Nombre de faux rejets}}{\text{Nombre total d'accès utilisateurs légitimes}} \quad (2.2)$$

A partir de ces deux taux, un troisième taux est calculé qui constitue un point de référence pour le choix du seuil de décision, c'est le taux d'équivalence des erreurs EER (*Equal Error Rate*) déterminé par le point d'intersection entre la courbe du taux de fausses acceptations et la courbe du taux de faux rejets en fonction du seuil de décision (figure 2.7). Idéalement, le système devrait avoir un EER égal à zéro. Comme ce n'est jamais le cas en pratique et puisqu'il est impossible de diminuer les deux erreurs en même temps (la variation des taux d'erreur FAR et FRR en fonction du seuil est inversée), il faut choisir un compromis entre FAR et FRR en ajustant la valeur du seuil de décision pour adapter le système avec le niveau de sécurité souhaité. En effet, pour une application qui demande une haute sécurité, on fixe le seuil de décision pour avoir le moins de FAR au détriment de FRR. En revanche, pour des applications personnelles, on privilégie un FRR bas.

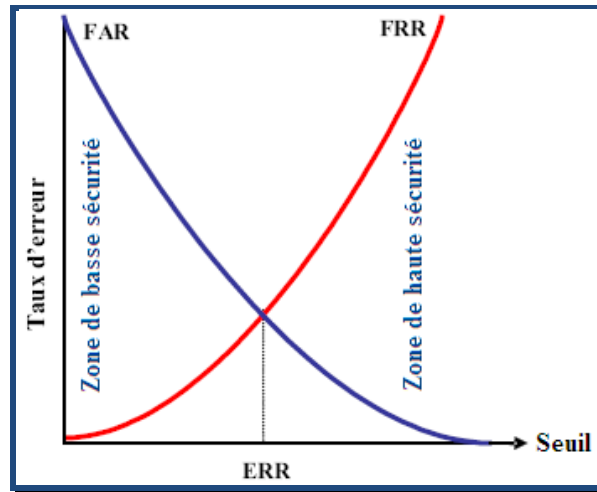


Fig. 2.7. Détermination du point d'équivalence des erreurs (EER)

Une autre courbe qui peut refléter les performances d'un système de vérification d'identité et la plus utilisée, c'est la courbe **ROC** (*Receiver Operating Characteristic* ou *Relative Operating Characteristic*) [6] [8]. Cette dernière comme il est illustré dans la figure 2.8, présente la variation du taux de vérification correcte (100-FAR (%)) en fonction du FRR.

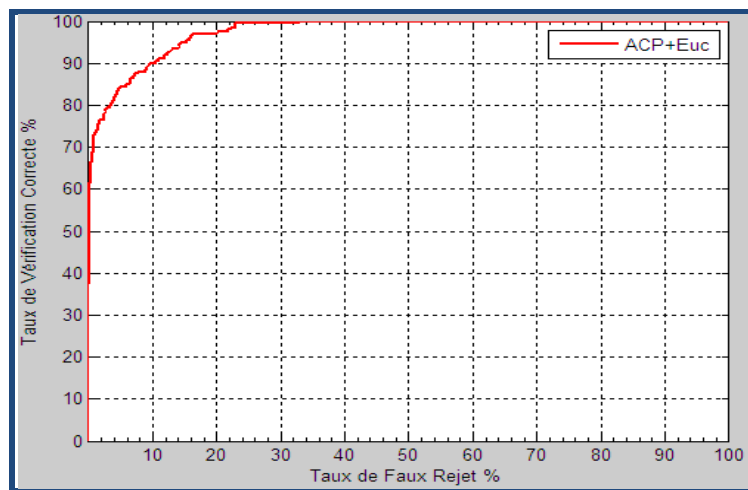


Fig. 2.8. Illustration de la courbe ROC

Cette courbe est fréquemment utilisée car elle permet de donner un aperçu sur la performance du système. En effet, plus cette courbe tend à épouser la forme du repère, plus le système est performant, c'est-à-dire possédant un taux de reconnaissance global élevé.

2.6.2. Le mode d'identification

Dans le mode identification, le système doit assigner à une personne à identifier l'identité offrant le score de similitude le plus élevé parmi toutes les mesures de similarité faites sur la base de référence. Dans ce scénario, un seul type d'erreur peut survenir à savoir la fausse identité, c'est dans le cas où le système assigne une fausse identité à une personne qui fait

partie ou pas de la base de données. De là, on peut évaluer un système d'identification de visage par le taux des fausses identifications FMR (*False Match Rate*) ou le taux des bonnes identifications (1-FMR) appelé généralement le taux de reconnaissance (*Recognition Rate*).

Le système qui a le taux de reconnaissance le plus grand est considéré le meilleur. Ce taux est la mesure la plus couramment utilisée pour l'évaluation d'un système d'identification de visage, mais il n'est pas toujours suffisant. En effet, en cas d'erreur, il peut être utile de savoir si le bon choix se trouve parmi les k premières réponses du système. Dans cette situation, l'évaluation peut se faire en utilisant la courbe de rang CMC (*Cumulative Match Characteristic*) [6] [8] qui représente la variation du taux de reconnaissance en fonction d'une variable que l'on appelle le rang. Ce dernier est le nombre k de décisions (k premières réponses) sur lequel est mesuré le taux de reconnaissance, avec k variant de 1 à M ; M étant inférieur ou égal au nombre de personnes enregistrées dans la base de référence. On dit qu'un système reconnaît au rang 1 lorsqu'il choisit la plus proche image comme résultat de la reconnaissance. On dit qu'un système reconnaît au rang 2, lorsqu'il choisit, parmi deux images, celle qui correspond le mieux à l'image d'entrée, etc. La figure 2.9 montre un exemple de courbe de rang. Plusieurs interprétations de ces courbes sont possibles :

- Le système qui obtient le meilleur taux pour des rangs faibles est considéré comme le meilleur. On dit que ce système atteint sa meilleure performance le plus rapidement possible.
- Plus le rang augmente, plus le taux de reconnaissance correspondant est lié à un niveau de sécurité faible.
- Pour comparer deux systèmes différents, il faut fixer un rang k inférieur ou égal à M .

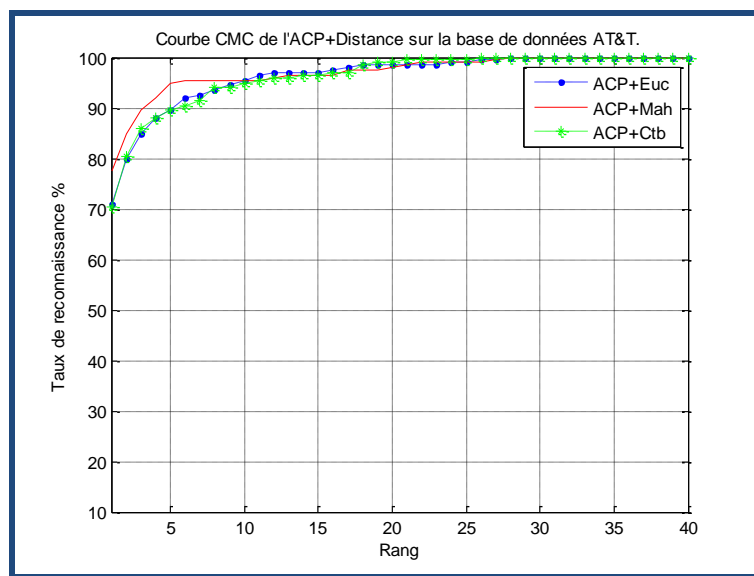


Fig. 2.9. Exemples de courbe CMC de la méthode ACP

2.7. Bases de données des images de visage

Plusieurs bases de données contenant des informations qui permettent l'évaluation des systèmes de reconnaissance de visages sont disponibles sur le marché. Toutefois, ces bases de données sont généralement adaptées aux besoins de quelques algorithmes spécifiques de reconnaissance ou elles sont collectées dans le cadre d'un projet. Ces bases de données de visage sont caractérisées par certains critères comme l'expression faciale, l'éclairage ou luminosité, l'occlusion, la variation de pose, le zoom sur le visage ou l'échelle, l'intervalle de temps entre les différentes sessions de prise de photos pour le même individu. Les paragraphes suivants récapitulent les principales caractéristiques des bases de données les plus célèbres.

2.7.1. La base de données FERET

La base de données FERET⁴ ("Face Recognition Technology") a été développée dans le cadre du programme FERET. Comme nous l'avons signalé dans le chapitre 1, Cette base a été recueillie et mise à disposition afin d'évaluer les progrès en reconnaissance automatique de visages. La collection de cette base de données est un effort de collaboration entre le Dr Wechsler et le Dr Phillips à l'Université George Mason. La première version de cette base a été rassemblée en 15 sessions, entre décembre 1993 et août 1996. Ensuite, la version 2 a vu le jour en 2003 avec des images couleurs de haute résolution (24-bit). Cette base est considérée parmi les plus anciennes et les plus volumineuses, et c'est probablement la base la plus connue et la plus utilisée par la communauté de reconnaissance de visages. Elle est actuellement administrée par le NIST (*National Institute of Standards and Technology*) aux États-Unis.

Les détails des caractéristiques de cette base seront présentés dans le chapitre 7 puisque nous avons utilisé cette base dans les tests expérimentaux.

2.7.2. La base AT&T

La base de données de visages AT&T⁵ (anciennement nommée ORL Database of Faces) a été développée dans le cadre d'un projet de reconnaissance de visages au sein du centre SVRG (Speech, Vision and Robotics Group) de l'Université de Cambridge. De même, nous détaillerons de plus la présentation de cette base dans le chapitre 7.

4. www.nist.gov/itl/iad/ig/colorferet.cfm

5. www.cl.cam.ac.uk/Research/DTG/attarchive/pub/data/att_faces.zip

2.7.3. La Base AR

La base de données AR⁶ a été créée par Aleix Martinez et Robert Benavente au Computer Vision Center (CVC) à l'UAB. Elle contient plus de 4000 images couleur correspondant aux visages de 126 personnes (70 hommes et 56 femmes). Chaque personne a participé à deux sessions espacées de deux semaines (14 jours). Les mêmes images ont été prises dans les deux sessions. Dans chaque session et pour la même personne, 13 images sont acquises en vue de face avec des conditions différentes d'expression faciale, d'éclairage, et d'occultation (lunettes de soleil et écharpe). Chaque personne a donc au total 26 images. Cette base est souvent utilisée pour évaluer des algorithmes de reconnaissance de visages sous des conditions de variation d'occultation.



Fig. 2.10. Extrait de la base de données de visage AR

2.7.4. La Base Extended Yale B (B+)

La base Extended Yale B (B+)⁷ est l'étendue de Yale Database Face B créée par l'université de Yale. Cette base est considérée comme la base standard pour l'évaluation de la robustesse des systèmes de biométrie faciale en cas d'illumination variable. Elle se compose de 16 128 images faciales de 10 individus capturées sous 9 poses et 64 conditions différentes d'éclairage (en total 576 images par personne). Le format des données de cette base de données est le même que la base de données de Yale Face B : format GIF en niveau de gris.

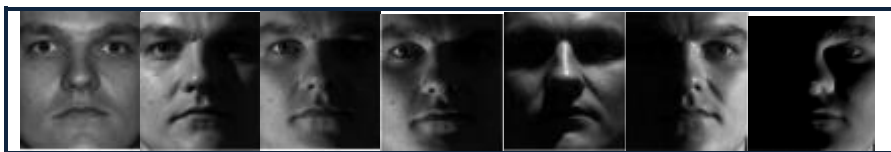


Fig. 2.11. Extrait de la base de données de visage Extended Yale B (B+)

2.7.5. La base Multi-PIE Face Database

La base de données de visage Multi-PIE⁸ recueillie à l'Université Carnegie Mellon en 2000, a largement contribué dans l'avancement de la recherche en reconnaissance de visage,

6 . www2.ece.ohio-state.edu/~aleix/ARdatabase.html

7 . www.vision.ucsd.edu/~iskwak/ExtYaleDatabase/ExtYaleB.html

8 . www.multipie.org/http://www.face-rec.org/databases/

surtout pour l'évaluation des algorithmes sous la contrainte de variation de pose et/ou d'éclairage. La base de données de visage Multi-PIE contient plus de 750 000 images de 337 personnes enregistrées dans un maximum de quatre sessions au cours d'une durée de cinq mois. Les images des sujets ont été capturées avec 15 points de vue et 19 conditions d'éclairage tout en affichant une gamme d'expressions faciales. Ces images d'une même session ont été capturées en quelques minutes les uns des autres en utilisant un système multi-caméra (15 caméras : Treize caméras étaient situées à hauteur de la tête, espacées à intervalles de 15° et deux caméras supplémentaires ont été situées au-dessus du sujet, simulant une vue typique de surveillance) et enregistrement multi-flash (18 flash). En outre, les images frontales à haute résolution ont été acquises ainsi. Au total, la base de données contient plus de 305 Go de données de visage.

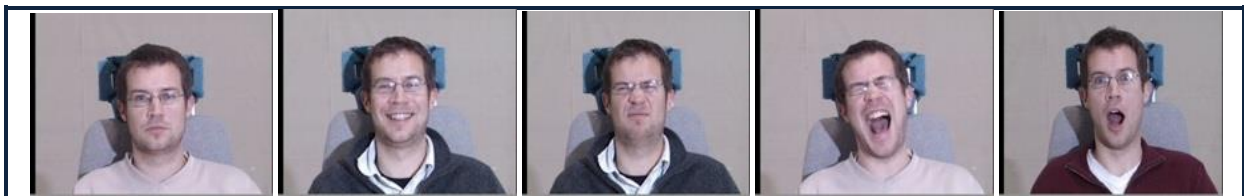


Fig. 2.12. Extrait de la base de données de visage Multi-PIE

Pour plus d'information sur les autres bases de données de visages, le lecteur peut se référer à la page web⁹.

2.8. Conclusion

Dans ce chapitre nous avons mis en relief les notions et les définitions de base liées à la biométrie de visage, ainsi que ses domaines d'application. Ensuite, nous avons décrit en détail les étapes indispensables au fonctionnement d'un processus automatique de reconnaissance de visages, allant de la capture de l'image à la prise de décision. Cette description nous a permis d'avoir une perspective correcte sur le système que nous désirons concevoir. Ainsi, nous avons mis en évidence les challenges liés à la reconnaissance de visages pour bien cerner les problèmes qui peuvent dégrader les performances d'un tel système dans des conditions d'acquisition difficiles et de proposer par la suite des solutions adaptées. Et finalement, une présentation des outils standards d'évaluation qui seront exploités dans la fin de cette thèse pour valider les performances de notre système développé. Dans le chapitre suivant, nous évoquerons l'état de l'art en reconnaissance de visages.

9 . www.face-rec.org/databases/

Chapitre 3

Etat de l'art en reconnaissance de visages

3.1. Introduction

Les premières études théoriques dans la reconnaissance automatique de visages, remontent au début des années 1960 lorsque Bledsoe a proposé dans son travail innovateur [25] le premier système semi-automatique pour la reconnaissance de visages. Dans ce système l'administrateur devait tout d'abord repérer manuellement des points caractéristiques sur l'image de visage tels que le nez, la bouche, les yeux et les oreilles, puis calculer les distances de ces points par rapport à un point de référence commun. Ces distances sont ensuite comparées à des données de référence dans une base de données. L'handicap de cette solution est la localisation manuelle des positions de ces points caractéristiques. Pour résoudre ce problème, en 1973 Kanade a mis en œuvre le premier système de reconnaissance faciale entièrement automatique dans son travail [26]. Mais le vrai déclenchement de la reconnaissance du visage a vu le jour en 1990, quand Kirby et Sirovich [27] ont appliqué une technique d'algèbre linéaire appelée analyse en composantes principales (ACP) pour modéliser un visage humain par un vecteur de dimension très réduite. Cette technique popularisée ensuite sous le nom « Eigenfaces » après l'amélioration apportée par Turk et Pentland [28], a donné une impulsion décisive à la communauté des chercheurs dans le domaine de la reconnaissance de visages. Ceci a permis de développer plusieurs approches qui exploitent divers techniques du traitement d'image (filtrage, transformation, opérateurs morphologiques, etc...) et les techniques d'analyse et de classification de données.

La grande majorité de la recherche et les produits commerciaux de la reconnaissance de visages agissent sur des images de visage à deux dimensions 2D de couleurs, d'intensité qui permettent d'obtenir la texture du visage et plus récemment des images prises avec des capteurs thermiques ou infrarouges. Cependant, vu qu'elles sont basées sur des données de nature pixel, ces méthodes ont montré leur insuffisance en robustesse devant les différents facteurs qui changent l'apparence du visage humain dans une image 2D (les facteurs déjà évoqués dans le chapitre 2). Pour surmonter ces problèmes, de nombreux chercheurs ont pensé à une autre représentation du modèle de visage qui ne sera pas influencée par la variation d'illumination, ni de pose. Cette représentation n'était que l'analyse d'image de visage en trois dimensions (3D) représentant à la fois le visage et la forme de la tête comme des données de profondeur qui redessine la structure géométrique du visage

Dans cette alternative, le modèle du visage est déterminé à partir des images 2.5D ou 3D (voir figure 3.1). L'image 2.5D (image de profondeur) correspond à une représentation

bidimensionnelle d'un ensemble de points 3D (x,y,z) où chaque pixel dans le plan X-Y stocke la valeur de sa profondeur z . On peut assimiler une image 2.5D à une image en niveau de gris où les pixels noirs correspondent au fond tandis que les pixels blancs représentent les points de surface les plus proches de la caméra. Par ailleurs, la méthode la plus simple pour représenter un visage 3D est le maillage polygonal 3D, ce dernier correspond à une liste de points connectés par des arêtes (polygones).

Dans la littérature, Il existe plusieurs techniques [29] [30] [31] [32] pour construire des modèles 3D complets d'un visage, les plus utilisées combinent des images 2.5D ou bien exploitent des systèmes d'acquisition 3D tel que le scanner 3D, un système multi-caméras (stéréoscopique), la lumière structurée ou les caméras de profondeur.

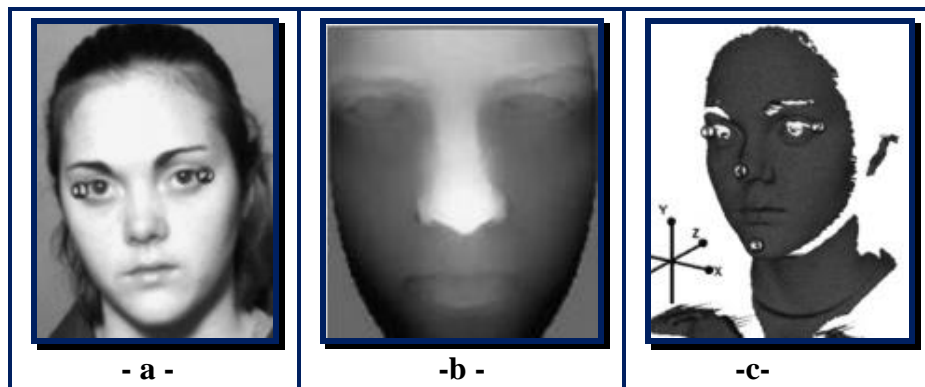


Fig. 3.1. Trois types de représentation d'une image de visage : a) image 2D, b) image 2.5D (image de profondeur) et c) image 3D (source: [33])

La reconnaissance 3D de visages semble être un outil prometteur surtout depuis l'apparition des dispositifs d'acquisition 3D performants. Son avantage principal réside dans le fait que la représentation 3D conserve toutes les informations sur la géométrie de visage, ce qui permet d'avoir une représentation réelle et plus précise des traits du visage, d'où on peut déceler des descripteurs plus puissants dans la discrimination des visages. Cependant, ces approches peuvent demander un coup de déploiement élevé dû à l'investissement des scanners 3D coûteux. Un autre inconvénient majeur concerne le grand volume de données tridimensionnelles qu'il est souvent nécessaire de convertir afin de pouvoir les traiter convenablement, ce qui implique une utilisation encore relativement inadaptée à des contraintes temps-réel, contrairement au traitement des images 2D.

Un troisième type d'approches de reconnaissance de visages pourrait être les algorithmes multimodaux 2D-3D qui combinent des informations provenant de l'image 2D ainsi que du modèle 3D de visage. L'avantage de cette catégorie est qu'il ajoute des informations de

profondeur aux approches classiques 2D sans trop augmenter le coût de calcul. Ces méthodes multimodales rapportent généralement des meilleurs taux de reconnaissance [33] [34].

Notre étude de l'état de l'art en reconnaissance de visages décrite dans ce chapitre se concentre spécifiquement sur les méthodes 2D, vu que notre contribution dans ce contexte utilise le modèle de visage 2D. Cependant, on pourra se référer aux travaux [33] [35] [36] pour plus de détails sur les techniques 3D et multimodales 2D-3D. En plus, la littérature sur la reconnaissance de visages 2D est vaste et diversifiée et la citation de la totalité de son état de l'art ne peut être complète. C'est pour cette raison que nous nous sommes donc intéressé à décrire la plupart des grandes méthodes classiques et la majorité des approches qui ont apporté une vraie originalité et une avancée sensible dans le domaine. Et pour avoir une idée sur les autres techniques non traitées ici, on peut consulter les résumés de l'état de l'art [37] [38] [39] [40] [41] [42].

Les approches 2D existantes peuvent être grossièrement divisées en trois familles : les approches globales, les approches locales, ainsi que les approches hybrides. Cette classification est la plus fréquente dans la littérature, mais elle n'englobe pas les autres techniques qui ne répondent pas aux critères de cette catégorisation.

3.2. Méthodes globales

Ces méthodes sont souvent appelées méthodes globales ou holistiques car elles mettent en valeur l'apparence globale du visage : le visage est traité comme un tout. Il n'est pas nécessaire de repérer certains points caractéristiques du visage. Ces méthodes visent généralement à réduire l'espace de représentation du visage par des techniques d'analyse statistique bien connues. C'est-à-dire, passer de la représentation du visage par une matrice de valeurs de pixels de taille très grande à un vecteur de taille très réduite. Ce passage est obtenu par la projection de l'image du visage, préalablement vectorisée, dans un nouvel sous-espace de plus faible dimension. Ce dernier est souvent conçu pour n'extraire de l'image de visage originale très redondante que les informations pertinentes, en générant en sortie un vecteur de taille très réduite et en même temps capable d'avoir des taux de reconnaissance raisonnables. C'est cette capacité qui permet de distinguer les techniques de modélisation par réduction d'espace présentées dans les paragraphes suivants.

3.2.1. La méthode Eigenfaces ou ACP

La méthode *Eigenfaces* (visages propres) est l'une des méthodes les plus largement utilisées pour la représentation du visage dans son ensemble. Elle se base dans son principe

sur la méthode mathématique d'analyse de données aléatoires : l'analyse en composantes principales (ACP) (*PCA, Principal Component Analysis*). L'ACP, également connue sous le nom de transformation de Karhunen-Loeve, est une méthode de projection linéaire qui permet de réduire la dimension de l'espace de représentation d'une distribution de données aléatoires en conservant ses composantes porteuses des informations les plus pertinentes et en ignorant celles les moins informatives. Cette technique a été utilisée au départ afin de représenter efficacement des images de visage humain dans le travail [27]. En 1991, cette technique a été reprise dans le cadre plus spécifique de la reconnaissance de visages sous le nom de la méthode des *Eigenfaces* [28]. Dans ces deux travaux, pour utiliser l'ACP sur une base d'apprentissage d'images de visage, chaque image à deux dimensions d'un visage est transformée en un vecteur obtenu en enchaînant les lignes (ou colonnes) de l'image. Ensuite, on détermine les vecteurs propres de la matrice de covariance formée par l'ensemble des vecteurs résultants. Les principaux vecteurs propres ou les visages propres sont les axes principaux du nouvel espace. La figure 3.2 montre les 17 premiers visages propres obtenus suite à l'application de l'ACP sur la base AT&T. Les coefficients de projection du vecteur associé à chaque image de visage dans ce nouvel espace fournissent un vecteur caractéristique (signature) de ce visage utilisé par la suite dans la classification ou la comparaison. Cette méthode sera présentée avec davantage de détails dans le chapitre 5.



Fig. 3.2. Les 17 premiers visages propres (eigenfaces) de la base AT&T

3.2.2. Variantes de la méthode ACP

Pendant les trois dernières décennies, plusieurs chercheurs ont développé diverses variantes et extensions de la méthode *Eigenfaces* (ACP), comme l'Analyse en Composantes Principales Bidimensionnelle (ACP2D) (*Two-Dimensional PCA*). Cette technique introduite par Yang et Zhang [43] traite l'image de visage comme une matrice plutôt qu'un vecteur en évitant l'étape de transformation de l'image en vecteur indispensable dans la méthode *Eigenfaces*. Les auteurs utilisent pour cela une matrice de covariance généralisée, calculée à

partir des lignes de l'image originale. Cette technique validée par [43] et [44], a prouvé sa robustesse vis-à-vis des différentes sources de variabilité que la méthode *Eigenfaces* et son efficacité pour l'optimisation du temps de calcul.

Une autre déclinaison de la méthode ACP nommée $(PC)^2A$ est présentée dans le travail [45]. Cette méthode qui vise l'enrichissement de l'information sur l'espace d'image original, effectuent deux projections (horizontale et verticale) de l'image de visage avant d'appliquer l'ACP. En effet, ces deux projections reflètent la distribution des caractéristiques faciales saillantes utiles pour l'identification de visage. L'image ou la carte de projection obtenue à partir de ces deux projections combinée avec l'image originale génère une image plus riche en information.

L'aspect de linéarité de la méthode *Eigenfaces* (ACP) peut être la cause d'une fausse classification car elle se base sur l'hypothèse que les variables (les visages) sont linéairement séparables dans l'espace de représentation. Or ce n'est généralement pas le cas. Par conséquent les deux approches de l'ACP et ACP2D ont été étendues à la version non linéaire appelées ACP à Noyaux (*Kernel Principal Component Analysis, KPCA*) [46] et ACP2D à Noyaux [47] en introduisant divers fonctions à noyaux non linéaires (kernel). L'idée est d'appliquer l'ACP à un nouvel espace de représentation des visages différent de celui des visages d'entrée, dans lequel les classes des visages sont censées être linéairement séparables. Cet espace est construit par une transformation non linéaire de l'espace d'entrée par des fonctions à noyaux non linéaires.

L'utilisation de ces méthodes de projection non linéaires permet dans une certaine mesure de réduire la dimension des images de meilleure façon et l'amélioration des taux de reconnaissance sur des jeux de tests donnés. Cependant, ces méthodes sont trop flexibles pour être robustes à de nouvelles données, contrairement aux méthodes linéaires.

Les méthodes que nous avons décrites jusqu'à maintenant, sont sans doute les plus connues des approches utilisant l'analyse en composantes principales (ACP) dans le contexte de la reconnaissance de visages. Mais, dans la littérature, de nombreuses méthodes génériques de l'ACP ont été proposées telles que : ACP Symétrique (*Symmetrical Principal Component Analysis*) [48], ACP Modulaire (*weighted modular principal component analysis*) [49] [50], ACP Pondéré Adaptative (*Adaptively weighted subpattern PCA*) [51] et Eigenfaces probabiliste [52].

3.2.3. La méthode Fisherfaces ou ADL de Fisher

L'ACP permet de réduire la dimensionnalité de l'espace de représentation de l'image de visage tout en optimisant sa reconstruction, mais elle n'optimise pas le facteur de discrimination entre les classes de visage. Pour résoudre ce problème, Belhumeur, Hespanha et Kriegman ont exploité dans leur travail [53] la technique d'Analyse Discriminante Linéaire de Fischer (ADL de Fisher) bien connue en anglais (*Fischer Linear Discriminant Analysis, FLDA*) pour donner naissance à une autre approche de la reconnaissance de visages souvent nommée *Fisherfaces*. Ce nom vient du fait que cette technique cherche à maximiser le critère mathématique de *Fisher*, à savoir le quotient de la variance inter-classe par la variance intra-classe. Autrement dit, L'ADL de Fischer cherche à déterminer un espace de projection (Espace des *FisherFaces*) qui permet de minimiser les variations entre les visages du même individu (intra-classe) et maximiser les variations entre les visages des individus différents (inter-classes) dans le but d'assurer une bonne séparation entre classes.

L'approche *Fisherfaces* possède des performances supérieures à celles qui sont données par la méthode *Eigenfaces* [38]. Cependant, ceci n'est plus vrai si un seul exemple d'apprentissage par personne est disponible, c'est-à-dire si les variations intra-classe sont nulles, alors les performances de *Fisherfaces* deviennent faibles par rapport à celles qui sont données par l'*Eigenfaces* [54]. Afin de remédier à ce problème, dans le travail de [55], les auteurs ont proposé de remplacer la matrice de dispersion intra-classe par une matrice constante. Ainsi, la méthode basée ADL se réduit alors à la méthode *Eigenfaces*.

Un autre problème de l'approche *Fisherfaces* est connu sous le nom de *Small sample size problem*. Etant donné que le nombre d'images est souvent inférieur à leur dimension, la matrice de variance intra-classe peut être singulière, et son inversion pose donc un problème. Des méthodes ont été proposées pour contourner ce problème, telle que le travail [56].

Le cadre théorique de la méthode ACP ou *Fisherfaces* sera présenté dans le chapitre 5.

3.2.4. Variantes de la méthode ADL de Fisher

De même que l'ACP, l'ADL de Fisher a été généralisée pour le cas non linéaire en utilisant des fonctions à noyaux non linéaires, à savoir l'analyse discriminante de Fisher à noyaux (*KFDA, Kernel Fisher Discriminant Analysis*) [57] et l'analyse discriminante directe à noyaux (*KDDA, Kernel Direct Discriminant Analysis*) [58].

Autres variantes de l'ADL de Fisher ont été développées afin d'améliorer la performance de cette approche. Parmi elles, l'analyse discriminante bilinéaire proposée par Visani et

Garcia [59]. Les auteurs de cette approche qui ont inspiré son concept à partir de l'ACP2D, ont rapporté sa supériorité en terme de taux de reconnaissance devant les méthodes : *Eigenfaces*, *Fisherfaces* et ACP2D. Ainsi, on trouve dans la littérature les variantes suivantes: [60] [61] [62] [63].

3.2.5. Analyse en Composantes Indépendantes (ACI)

L'analyse en composantes indépendantes (ACI) (*ICA, Independent Component Analysis*) est un outil d'analyse de données aléatoires permettant de trouver une nouvelle base de représentation dans laquelle les projections des observations sur cette base fournissent des composantes indépendantes. L'ACI dans son contexte mathématique, est une généralisation de l'ACP qui utilise en plus des statistiques d'ordre 2, les statistiques d'ordre supérieur (*HOS, High Order Statistics*), ce qui peut produire une représentation plus puissante. Ainsi, les composantes recherchées dans la nouvelle base ne sont pas forcément orthogonales, alors qu'elles le sont pour l'ACP. Et dans le contexte physique, elle peut être assimilée à un problème de séparation de sources comme initialement formulé dans [64] dans le sens où elle permet d'extraire les structures fondamentales (les sources) d'une donnée (signal, image, ...).

L'analyse en composantes indépendantes (ACI) a été introduite par les spécialistes du traitement de signal pour analyser des signaux issus de multiples capteurs. Outre le traitement de signal, cette technique a été utilisée dans d'autres domaines, en l'occurrence la reconnaissance de visages [65] [66] [67]. Le principe de base de l'ACI dans ce cadre, consiste à considérer la matrice des observations (les images de visage) comme étant une combinaison linéaire et indépendante des sources. Le but de l'ACI est donc de trouver à partir des images de visages références une estimation de la matrice des sources. Ces dernières sont des images de base statistiquement indépendantes. Ensuite, les coefficients de la combinaison linéaire de ces images de base qui forme chaque image de visage, constituent son vecteur caractéristique.

Les auteurs des articles [66] [67] ont proposé deux architectures différentes pour exprimer la matrice des images de visage. La première architecture considère que les images sont des variables aléatoires alors que la seconde architecture considère les pixels comme étant des variables aléatoires. L'application de ces deux architectures permet d'obtenir deux types d'images de base (sources) différentes. La figure 3.3 illustre les résultats obtenus après l'application de l'ACI avec ces deux architectures.

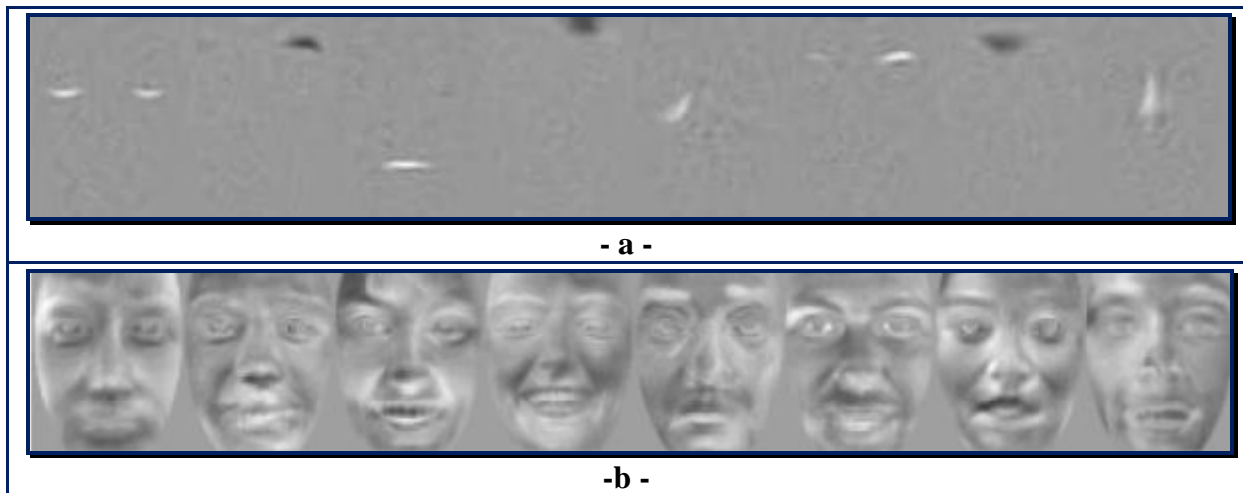


Fig. 3.3. Exemples des images de base obtenues avec la méthode ACI: a) pour l'architecture 1 et b) pour l'architecture 2. (Source [67]).

De nombreuses variantes de l'ACI ont été proposées pour apporter des améliorations. On trouve en particulier :

- La version non linéaire appelée analyse en composantes indépendantes à noyaux (*KICA, Kernel Independent Component Analysis*) [68] qui cherche à résoudre le problème de séparation des observations par la version non-linéaire.
- L'ACI rapide (*Fast ICA, Fast Independent Component Analysis*) [69] [70] qui a permis de remédier le problème du temps de calcul observé avec l'ACI classique.
- La combinaison entre l'analyse en composantes indépendantes et l'analyse discriminantes de Fisher [71].
- L'ACI par blocs (*Block independent component analysis*) [72] qui subdivise l'image de visage en ensembles de blocs et applique l'ACI à chacun.

En termes de comparaison, il a été démontré que l'ACI, ainsi que ses différentes variantes, est très performante par rapport à l'ACP [73].

3.2.6. Transformée de cosinus discrète (DCT)

La transformée de cosinus discrète (*DCT, Discrete Cosine Transform*) introduite au début des années soixante-dix dans le travail [74], est fréquemment utilisée dans le domaine de compression multimédia. Mais, elle était une voie très prometteuse dans la reconnaissance de visage [75] [76] [77] [78] [79] parce qu'elle a permis de minimiser les effets de variation de pose, de taille et d'alignement des visages. Cette méthode est beaucoup plus rapide et simple concernant l'extraction du vecteur caractéristique d'un visage si on le compare avec l'ACP et l'ADL. Un autre avantage de cette méthode repose sur sa grande flexibilité en cas d'ajout des

images de visages puisque cette opération ne demande aucun réapprentissage complet, contrairement aux méthodes de projection linéaires ou non linéaires. En effet, dans le cas de la DCT, chaque image de visage est représentée par un vecteur composé des premiers coefficients de sa transformée DCT indépendamment des autres images.

3.3. Méthodes locales

Les méthodes locales essaient de modéliser le visage en s'approchant à la manière avec laquelle l'être humain perçoit la morphologie du visage, c'est à dire, au processus d'analyse de notre système visuel à nos traits de visage pour identifier une personne. Elles permettent de représenter un visage grâce à des caractéristiques multiples beaucoup plus robustes à certaines variations comme les variations d'illumination et de pose. Il existe deux stratégies pour cela: les méthodes basées sur des caractéristiques locales (*Local feature based method*) et les méthodes basées sur l'apparence locale du visage (*Local appearance based method*).

3.3.1. Méthodes basées sur des caractéristiques locales

On les appelle aussi les méthodes à traits locaux ou à caractéristiques d'intérêt car elles se basent en premier temps sur la détection des points d'intérêt dans le visage. Ces points sont généralement des points particuliers qui se situent sur les frontières des traits remarquables du visage tels que : les yeux, le nez, la bouche, les sourcils, etc... . Ensuite, un modèle de visage est construit à partir de ces points par deux alternatives différentes. La première, par l'extraction des mesures géométriques prises sur ces points, ce sont donc les méthodes géométriques. Tandis que pour la deuxième, par une fonction codée via un graphe créé à partir de ces points, ce sont les méthodes à graphe.

3.3.1.1. Méthodes géométriques

Les premières méthodes proposées pour la reconnaissance de visages par ordinateur appartiennent à cette catégorie [25] [26]. Elles s'appuient toutes sur l'extraction des mesures géométriques prises sur les points d'intérêts détectés comme il est déjà signalé. Parmi les mesures importantes, on trouve notamment : les positions et les largeurs de la bouche, des yeux, du nez et des sourcils, ainsi, les distances ou les angles entre ces points. Les mesures extraites sont ensuite rassemblées dans un vecteur (vecteur caractéristique) pour les utiliser dans la phase de classification.

Dans la littérature, il existe plusieurs techniques géométriques qui se distinguent entre elles par le nombre et les positions des points détectés et la nature des mesures prises sur ces

points. Les figures 3.4 (a à c) illustrent des exemples de la répartition des points détectés et les mesures choisies dans les approches géométriques [80] [81] [82].

Les comparaisons menées sur les mêmes données en opposant les méthodes géométriques à d'autres méthodes ont montré clairement leurs limitations et faiblesses. En effet, la performance de ce type de méthodes dépend de la phase de détection et d'extraction des points caractéristiques du visage. Or, cette tâche est difficile et dans certains cas, elle est impossible, en particulier dans les cas où des occultations ou des variations (pose, expression) de visages sont présentes. Face à ce problème majeur, les recherches sont orientées vers le développement des algorithmes de détection des points caractéristiques robustes et précis.

Malheureusement, bien que les performances de ces algorithmes de détection soient largement démontrées, les méthodes de reconnaissance de visages basées uniquement sur les caractéristiques géométriques restent insuffisamment fiables ou robustes. Cela est dû à la nature des informations extraites (les mesures géométriques) qui sont insuffisantes pour discriminer les visages entre eux parce qu'on trouve des distances qui sont constantes entre des différentes personnes, tandis que d'autres informations utiles telles que les valeurs des niveaux de gris autour des points caractéristiques sont complètement écartées. Pour résoudre ce problème, les méthodes basées sur les graphes ont vu le jour.

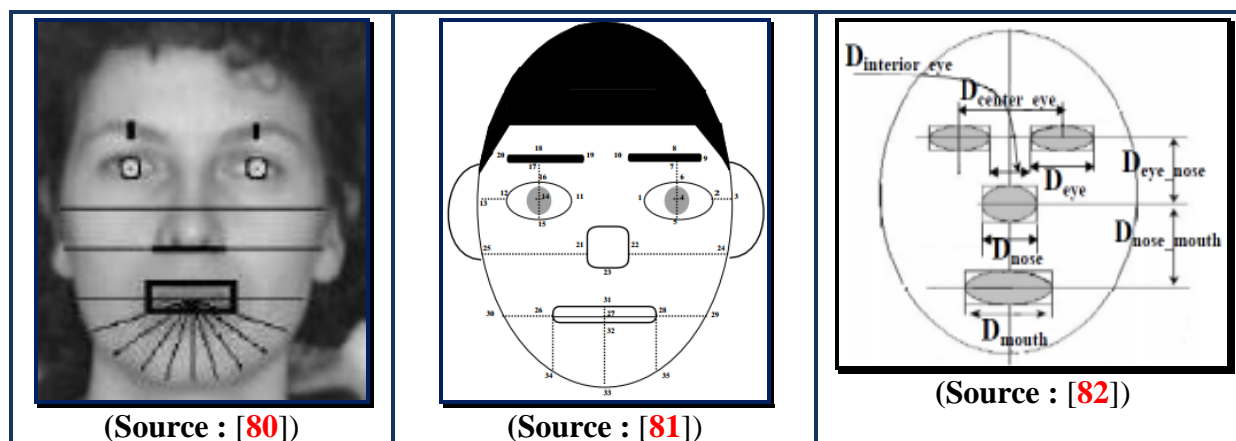


Fig. 3.4. La répartition des points détectés et les mesures choisies dans trois approches géométriques.

3.3.1.2. Méthodes basées sur les graphes

Ces méthodes représentent une image de visage à l'aide d'un graphe formé par un ensemble de nœuds reliés entre eux par des arêtes. Ces nœuds peuvent être des points d'une grille régulière virtuelle déposée sur le visage (figure 3.5) ou des points caractéristiques du visage (figure 3.6). Dans ces approches, on vise à modéliser le visage par deux types d'information. La première est de nature locale, extraite par une analyse de la texture du

voisinage de chaque nœud en utilisant des filtres d'image appropriés tel que le filtre de Gabor ou par analyse de sa forme par des opérateurs morphologiques. Tandis que la deuxième est de nature globale, extraite des positions relatives des nœuds.

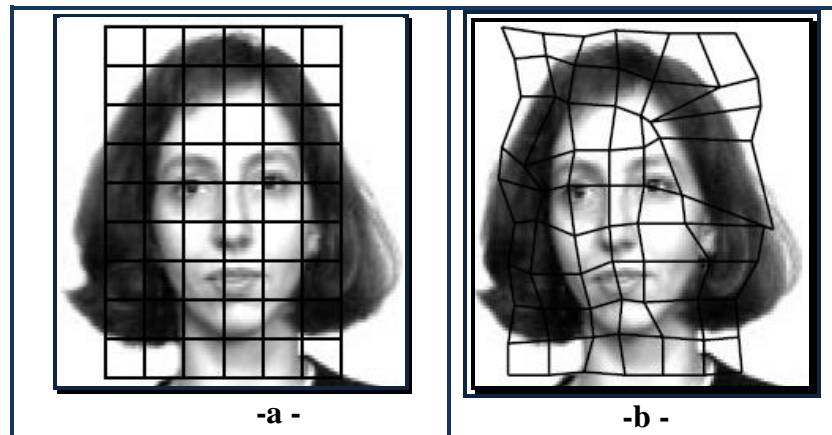


Fig. 3.5. Exemple de graphe élastique avec une topologie de grille : a) le graphe de l'image de visage de la base de données, b) le graphe de l'image de visage à reconnaître après son déformation. (Source [84])

La première approche développée dans ce sens est celle de Manjunath et al. [83]. Dans ce travail, les auteurs détectent des points caractéristiques à l'aide d'ondelettes de Gabor. Ensuite, un graphe fixe est construit à partir de ces points détectés. Une fois la topologie du graphe formulée, la reconnaissance se fait en comparant le graphe de l'image à reconnaître avec l'ensemble des graphes (ou modèles) des images de la base de données. Cette comparaison met en jeu la similarité entre les caractéristiques de chaque nœud extraites par analyse de filtre de Gabor et les similarités sur la topologie globale de chaque graphe, par comparaison des distances entre les nœuds du graphe. Les tests, faits par la suite sur cette approche, ont montré ses limites face aux variations des expressions et de pose du visage à cause de la topologie fixe du graphe. Pour palier ce problème et en inspirant de l'idée du graphe, Lades et al. ont proposé d'utiliser un graphe topologique élastique (déformable) dans leurs travaux [84]. Dans ce dernier, les auteurs utilisent un modèle pour la reconnaissance des objets appelé Architecture de Liens dynamiques (*DLA, Dynamic Link Architecture*), qu'ils ont appliqué aussi sur les visages en donnant naissance à un nouveau algorithme de reconnaissance de visages nommé *Elastic Graph Matching (EGM)*. Cet algorithme extrait le même type d'informations que [83] dans la phase d'apprentissage avec la différence qu'il utilise un graphe avec une topologie de type grille rectangulaire localisée sur l'image de visage (figure 3.6) où les nœuds sont étiquetés avec les réponses des filtres de Gabor dans plusieurs orientations et plusieurs fréquences spatiales appelées jets. Tandis que, pour la phase

de reconnaissance, le graphe représentatif de l'image de visage à reconnaître subit une déformation à chaque fois où on détermine sa correspondance avec un graphe des images de visage de la base de données. Cette déformation prend une position finale au moment où la similarité entre les jets des deux images prend la valeur maximum. L'image de visage de la base de données qui donne un coût de déformation le plus faible, est prise comme identité de l'image de test. Ce coût de déformation, comme il est indiqué dans [84] est une fonction qui code la déformation du graphe en prenant en compte la similarité entre les jets et la variation des distances entre les nœuds.

Plusieurs améliorations de la méthode *Elastic Graph Matching* ont été proposées. La plus populaire entre elles, celle de Wiskott et al. [85] [86] est connue sous le terme d'*Elastic Bunch Graph Matching (EBGM)*. Les auteurs de cette méthode ont utilisé cette fois-ci un graphe dont les nœuds sont situés sur un certain nombre de points caractéristiques du visage (figure 3.6.a). De plus, ils ont introduit la notion de *Face Bunch graphe (FBG)* dans le but d'utiliser un seul graphe pour toutes les images d'une même personne de la base. Un FBG est une représentation générale du visage pour laquelle on associe à chaque nœud, un ensemble de jets (Bunch) extraits à l'aide des filtres de Gabor représentant toutes les variabilités possibles de ce nœud (figure 3.6.b). Pendant la phase d'enrôlement, le visage de chaque personne sera modélisé par un FBG construit d'une manière supervisée, en combinant des informations à partir d'un certain nombre de graphes de ce visage. Alors que, dans de la phase de reconnaissance, un seul jet du FBG est retenu pour chaque nœud, celui qui maximise la fonction de similarité avec les jets du visage à reconnaître.

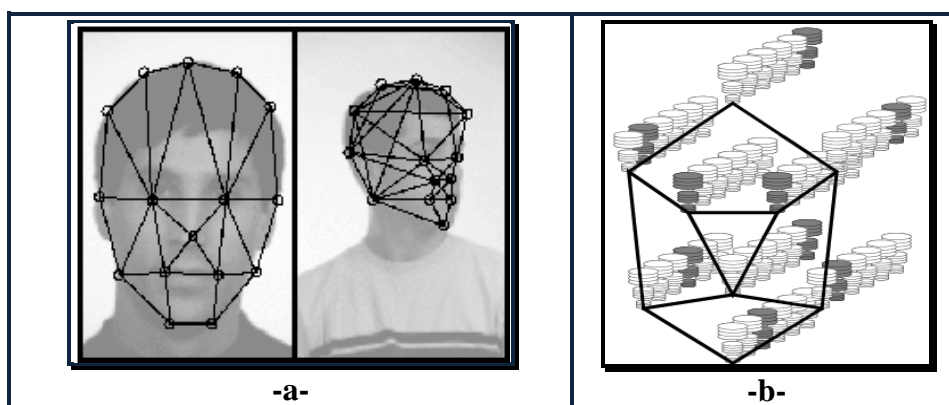


Fig. 3.6. Exemple de graphe élastique avec des points caractéristiques: a) le graphe d'un visage de la base de données et le graphe du visage à reconnaître après son déformation, b) Exemple de Face Bunch graphe (Source : [86]).

L'EBGM est classée parmi les méthodes les plus performantes en termes de robustesse. Cependant, elle a deux inconvénients qui ont été mis en évidence par [87] lors de sa

comparaison avec deux autres méthodes : *Eigenface* et les réseaux de neurones auto-associatifs. Premièrement, elle exige un temps de calcul supérieur aux autres méthodes ce qui la rend plus difficile à mettre en œuvre en pratique. Deuxièmement, seules les informations sur les positions clés de l'image sont utilisées pour l'identification.

Malgré les problèmes de l'EBGM, elle a été une source de motivation pour certains travaux qui ont proposé des variantes à cet algorithme, notamment celles qui utilisent d'autres descripteurs pour l'extraction des jets que les ondelettes de Gabor, comme le cas de l'approche proposée dans [88] très connue sous le nom *Morphological Elastic Graph Matching (MEGM)*, la correspondance élastique des graphes morphologiques. Les auteurs de cette approche utilisent des opérateurs morphologiques de dilatation et d'érosion pour extraire les coefficients des Jets. Les résultats obtenus par cette approche sont comparables à ceux de l'approche originale mais avec l'avantage qu'elle ne nécessite pas le calcul des banques des filtres de Gabor.

Une autre variante importante de l'EBGM a été développée dans les travaux [89] [90]. Les auteurs de ces derniers ont remplacé les filtres de Gabor par un histogramme des orientations de gradients (HOG) dans des fenêtres centrées sur les nœuds du graphe. Les résultats des expériences montrent un meilleur comportement de la performance par rapport à celle d'EBGM classique. Cette meilleure performance s'explique par les propriétés des descripteurs de HOG qui sont plus robustes aux variations de l'éclairage et de la rotation, et grâce à une plus grande précision des graphes de face.

3.3.2. Méthodes basées sur l'apparence locale du visage

Les performances des méthodes basées sur les caractéristiques locales chutent radicalement lorsqu'elles sont confrontées à des images à fortes variations d'expression, d'illumination ou des occultations partielles du visage car elles dépendent de la robustesse de l'algorithme de détection des points caractéristiques. Cette tâche de détection qui doit être précise, n'est pas facile et n'a pas été résolue de manière fiable. C'est pourquoi, d'autres approches qui ne nécessitant pas une localisation précise de ces points ont été proposées. Ce sont les méthodes basées sur l'apparence locale du visage (*Local appearance based method*). L'ensemble de ces méthodes ont en commun une première étape qui consiste dans le découpage de l'image de visage en régions (patches). Ces dernières peuvent avoir une forme rectangulaire, elliptique, etc. (figure 3.7) mais ce qui est le plus largement utilisé est le découpage rectangulaire. Les fenêtres peuvent être superposées ou non et la taille de la région a une influence directe sur le

nombre de caractéristiques et la robustesse de la méthode. Une fois que les régions locales ont été définies, l'étape clé de ces approches est abordée, c'est l'extraction des caractéristiques. Dans cette étape critique, il faut choisir la meilleure manière de représenter les informations de chaque région en introduisant un descripteur. Les vecteurs obtenus pour chaque région sont ensuite concaténés pour obtenir le vecteur caractéristique final.

Dans la littérature, plusieurs descripteurs sont exploités dans ce sens et chacun d'eux a montré son aptitude de robustesse face à un type de variation d'apparence donné :

- Les ondelettes de Gabor [91] [92] et les histogrammes *LBP (Local Binary Pattern)* [93] [94] [95] connues pour être relativement robustes aux variations d'illumination, d'expression et de pose.
- Le descripteur *POEM (Patterns of Oriented Edge Magnitudes)* [96] et les histogrammes des orientations de gradients (*HOG*) [97] qui sont robustes aux variations de pose.

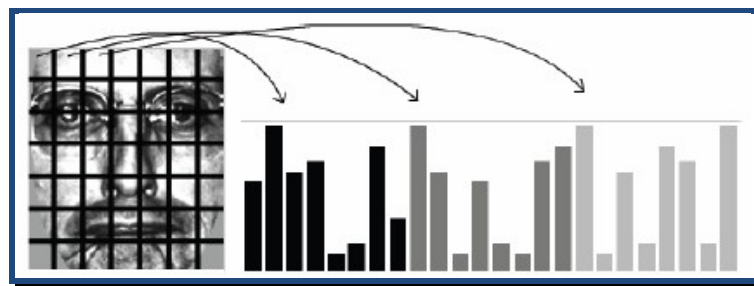


Fig. 3.7. Exemple de représentation d'un visage par la méthode d'apparence local basée sur les histogrammes LBP.

3.4. Méthodes hybrides

Les méthodes hybrides résultent de l'association (ou fusion) des méthodes locales et des méthodes globales. Elles cherchent à modéliser un système de reconnaissance de visages regroupant l'information sous forme locale et globale, dans le but de s'approcher du système visuel humain qui combine ces deux types d'informations pour reconnaître un visage. Les facteurs clés qui influent les performances de ces méthodes hybrides comprennent le choix des caractéristiques pour la combinaison et la manière de les combiner de telle sorte que leurs avantages soient préservés et que leurs inconvénients soient évités. Ces caractéristiques présentent en effet des propriétés différentes et complémentaires: alors que les caractéristiques locales sont très sensibles au bruit, les caractéristiques globales, quant à elles, le sont peu. En revanche, ces dernières sont très sensibles à la pose contrairement aux caractéristiques locales. C'est pour cette raison qu'il n'existe pas à ce jour un algorithme permettant de combiner de façon optimale ces informations. Néanmoins, les méthodes

hybrides développées jusqu'à présent, essaient uniquement de tirer partie des avantages des deux types de méthodes citées plus haut et proposer des solutions pour éviter leurs inconvénients. A titre d'exemples de ces méthodes, on trouve :

- L'approche appelée Analyse en Composantes Locales (*LCA, Local Component Analysis*) a été proposée dans le travail [98]. Plusieurs analyses en composantes principales sont réalisées pour extraire différentes caractéristiques locales. Celles-ci sont ensuite combinées et une procédure minimisant l'erreur de reconstruction avec une contrainte parcimonieuse permet de réaliser la reconnaissance.
- L'approche ACP-LogGabor (*PCA-LG, Principal Component Analysis-LogGabor*) [99] qui effectue une convolution avec des ondelettes de Gabor orientées autour de certains points caractéristiques du visage afin de créer des vecteurs contenant la localisation et la valeur d'amplitudes énergétiques locales. Ces vecteurs sont ensuite envoyés dans un algorithme ACP afin de réduire la dimension des données.
- Les deux récentes approches [100] [101] qui utilisent dans le même sens les ondelettes de Gabor et respectivement Kernel-ACP et ACP.

3.5. Comparaison

Pour finir ce chapitre, nous donnons une comparaison globale de ces méthodes.

Les méthodes globales présentent un certain nombre d'avantages :

- Elles sont rapides à mettre en œuvre. En effet, le problème de la reconnaissance faciale automatique est transformé en un problème d'analyse de sous-espaces de visages par méthodes statistiques reposant sur des calculs matriciels relativement simples
- Elles sont souvent applicables à des images basses résolution ou de mauvaises qualités.

Certains inconvénients apparaissent cependant avec ces méthodes globales :

- Elles ne sont robustes qu'à des variations limitées (pose, illumination, expression). Ceci se comprend aisément puisqu'elles considèrent le visage comme un tout. Et la moindre variation des conditions de l'environnement ambiant entraîne des changements inéluctables dans les valeurs des pixels qui sont traités directement.
- Il est nécessaire de disposer de suffisamment de données représentatives des visages.

De la même manière les méthodes locales présentent certains avantages :

- Le modèle créé possède des relations intrinsèques bien définies avec les visages réels parce qu'elles prennent en compte la particularité du visage en tant que forme naturelle à reconnaître.
- Les modèles créés peuvent prendre en compte explicitement les variations telles que la pose, l'illumination ou les expressions. La reconnaissance est ainsi plus efficace dans le cas de fortes variations.
- La connaissance a priori sur les visages peut être intégrée aux modèles afin d'améliorer leur efficacité.

Les méthodes locales présentent cependant quelques inconvénients :

- La construction du modèle, reposant souvent sur la détection de points caractéristiques faciaux, peut être laborieuse et lourde. En effet, il faut souvent placer manuellement un assez grand nombre de points sur le visage ou automatiquement à l'aide d'un algorithme de détection ayant une grande précision.
- L'extraction des points caractéristiques peut être difficile dans le cas de variations de pose, d'illumination, d'occlusion . . .
- Les images doivent être relativement de bonne qualité, et/ou être de résolution suffisante afin de pouvoir extraire les points caractéristiques.

3.6. Conclusion

Dans ce chapitre, nous avons présenté les grandes lignes de l'évolution de la reconnaissance de visages. Comme déjà évoqué dans l'introduction, ce chapitre n'a pas pour finalité de décrire tous les algorithmes de reconnaissance de visage mais nous nous sommes contentés de présenter les approches qui ont introduit une nette évolution dans le domaine de cette biométrie, permettant ainsi une véritable amélioration des performances. Nous avons décrit également les avantages et les inconvénients de ces approches pour situer les unes par rapport aux autres. Cette comparaison nous a permis de conclure qu'aucune méthode n'occupe le premier rang devant toutes les autres.

Au moment du développement d'une approche de reconnaissance de visage, il faut tout d'abord, dégager les problèmes liés à l'environnement réel de l'application envisagée. Par la suite, un compromis doit être réalisé entre son taux de reconnaissance et sa rapidité.

Dans le chapitre suivant, nous présentons l'approche que nous avons développée pour la détection et la localisation d'un visage dans une image fixe.

Chapitre 4

Détection et localisation de visages

4.1. Introduction

La détection de visage, cas particulier de la reconnaissance des formes, est un thème de recherche en plein essor dans les domaines du traitement d'image et de la vision assistée par ordinateur. La raison qui explique un tel effort, est la variété des champs d'application de la détection de visages. Ainsi, cette dernière est un processus très utilisé non uniquement dans la reconnaissance de visages, mais dans d'autres applications telles que: la reconnaissance des expressions faciales, le suivi de visage, l'indexation d'images, la visiophonie, les systèmes de téléconférence, etc. ...

Dans le cas d'un système de reconnaissance de visages, la détection de visages est requise comme étape cruciale qui vient en amont de l'extraction des caractéristiques. Elle consiste à détecter la présence de chaque visage humain dans une image numérique ou dans une séquence vidéo et localiser en conséquence son emplacement. Un tel système ne pourra jamais devenir intégralement automatique s'il n'a pas été précédé par une phase de détection automatique.

La qualité de l'algorithme de détection de visage a un impact direct sur la fiabilité et la robustesse du système entier de reconnaissance. En effet, les performances de cet algorithme sont souvent dégradées par la variabilité des conditions d'acquisition de l'image de visage déjà citée dans la section 2.5 du chapitre 2. En conséquence, un visage non détecté correctement (visage non ou partiellement détecté ou fausse détection d'un objet en tant que visage), entrainera une anomalie dans la suite du système (extraction des caractéristiques et classification). C'est pour cette raison que la détection de visage reste un verrou technologique non résolu parfaitement.

Dans ce chapitre, nous allons décrire les aspects théoriques et pratiques d'une nouvelle approche que nous avons développée pour la détection et la localisation de visages dans une image fixe à multiples visages. Mais, avant cette description, nous passons en revue les méthodes les plus connues proposées dans ce cadre. Et par l'occasion, nous nous focalisons sur une méthode considérée comme l'une des plus efficaces en détection de visage, la méthode de Viola et Jones [102] [103]. Cette méthode qui est devenue rapidement un standard dans ce domaine, sera pour nous une référence de comparaison de notre approche dans la partie expérimentation.

4.2. Méthodes de détection de visages

Dans la littérature, on trouve une multiplicité d'approches proposées par les chercheurs dans le but de surmonter les problèmes de la détection de visages et d'atteindre un niveau avancé. La plupart de ces approches peuvent être répertoriées en quatre catégories principales [104] [105] [106].

4.2.1. Approches basées sur les connaissances acquises

Ces approches essaient de s'approcher à la manière de perception visuelle de l'être humain à un visage, puisqu'elles exploitent les connaissances a priori que l'être humain a acquises sur la morphologie d'un visage typique, utilisées dans la vie quotidienne pour détecter un visage parmi plusieurs objets dans une scène. Par exemple, un visage apparaît souvent dans une scène avec deux yeux symétriques, un nez et une bouche. Ces connaissances sont codées par des critères descriptifs et des règles métriques définis à partir des rapports entre les caractéristiques faciales (le nez, la bouche et les yeux) ou les régions qui les entourent. Les règles les plus utilisées sont généralement inspirées:

- Des distances relatives et les positions des caractéristiques du visage (les yeux, la bouche et le nez).
- De la symétrie entre les yeux.
- De la différence entre la moyenne des valeurs d'intensité de la partie centrale et la partie supérieure du visage.

Dans ces approches, après avoir défini les règles, chaque région candidate de l'image d'entrée est soumise à un processus de vérification qui permet d'accepter cette région en tant que visage si elle vérifie ses règles ou non.

Parmi les approches proposées dans cette catégorie, on peut citer les travaux suivants : [107] [108] [109].

4.2.2. Approches basées sur le Template-matching

Ces méthodes utilisent des modèles standards (Template-matching) d'un visage qui décrivent le visage comme un tout ou séparément les traits du visage. Ces Templates peuvent être définies soit manuellement, soit paramétrées à l'aide des fonctions. Ensuite, la corrélation est calculée, entre l'image candidate de visage et la Template pour déterminer leur correspondance. Ces méthodes ont été utilisées à la fois pour la localisation et la détection de

visage [110] [111] [112], mais, elles rencontrent encore quelques problèmes de robustesse liés aux variations de lumière, d'échelle, etc.

4.2.3. Approches basées sur l'apparence

Ces approches traitent la détection comme un problème de classification (visage, non-visage). Elles utilisent en premier lieu une analyse statistique d'un ensemble d'images représentatives de la variabilité de l'aspect facial pour les représenter par des modèles. En deuxième lieu, elles appliquent des techniques d'apprentissage automatique de ces modèles pour construire des machines capables de séparer les visages des non-visages. Ainsi, chaque région extraite de l'image originale en tant que candidate de visage, subit la même analyse statistique et son modèle est soumis à l'entrée de la machine de classification. Plusieurs techniques ont été proposées dans cette catégorie, notamment, celles qui utilisent les réseaux de neurones telles que : [113] [114] [115] [116] et autres dont leur aperçu est donné dans l'enquête de la littérature [117] qui comprend un examen général de quelques systèmes de détection de visage basés sur les réseaux de neurones, les machines à vecteurs de support (SVM) [118] [119] [120] [121] et les algorithmes de classification AdaBoost et ses dérivés [102] [103] [122] [123] [124] [125] [126].

Méthode de Viola et Jones

A cette catégorie des approches basées sur l'apparence, appartient l'une des méthodes les plus efficaces et les plus célèbres en détection de visages, celle proposée par Viola et Jones [102] en 2001, puis révisée par les mêmes auteurs en 2004 [103]. Cette méthode modélise des milliers d'exemples de visage et non-visages par extraction des caractéristiques nommées « *pseudo-Haar* » (inspiré de l'extraction des caractéristiques par les ondelettes de *Haar*). Ces dernières sont calculées par la différence des sommes de pixels de deux ou plusieurs zones rectangulaires adjacentes. La figure 4.1 donne des exemples des caractéristiques proposées par Viola et Jones à 2, 3 ou 4 rectangles, dans lesquelles la somme de pixels sombres est soustraite de la somme des pixels blancs. Ensuite, les modèles calculés sont pris comme des stimulés pour entraîner l'apprentissage supervisé d'un classificateur. Ce dernier se base sur le principe de « *Boosting* » qui consiste à construire un classificateur fort (classificateur capable de produire des décisions très précises) à partir de la combinaison de plusieurs classificateurs faibles boostés en cascade (32 classificateurs dans la version originale). Un classificateur est dit faible dans le langage du « *Boosting* » lorsqu'il ne possède pas des bonnes performances

de classifications. Il doit seulement pouvoir classifier plus que 50% des données correctement.

En phase de détection, l'ensemble de l'image est parcouru en déplaçant la fenêtre de détection d'un certain pas dans le sens horizontal et vertical, avec changement d'échelle de la fenêtre à chaque nouveau balayage de l'image. Pour chaque position de cette fenêtre, on calcule un ensemble de caractéristiques « *pseudo-Haar* » utilisées par chaque classificateur faible, en commençant par le premier étage. Si la réponse d'un classificateur est négative, la région localisée par la fenêtre est jugée comme non-visage. Sinon, on passe au classificateur suivant. Enfin, si tous les étages répondent positivement, la région localisée est détectée en tant que visage.

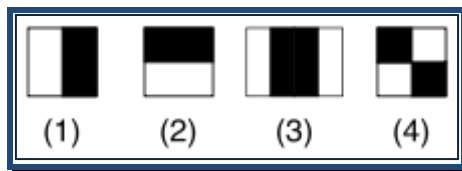


Fig. 4.1. Exemples de caractéristiques pseudo-Haar utilisées par Viola et Jones

De très nombreux travaux ont été publiés par la suite, visant à améliorer le paramétrage de la méthode, ou à en combler un certain nombre de ses limitations [122] [123] [124] [125] [126]. Les auteurs de ces derniers ont proposé d'étendre ou remplacer la technique de « *pseudo-Haar* » par d'autres techniques d'extraction des caractéristiques et/ou d'utiliser des variantes de l'algorithme de **Boosting** dans la phase de classification.

La méthode est par ailleurs reconnue comme étant celle ayant eu le plus d'impact dans le domaine de la détection de visage dans les années 2000.

4.2.4. Approches basées sur des caractéristiques invariantes

Contrairement aux approches fondées sur les connaissances acquises, les chercheurs ont essayé de trouver et d'exploiter des caractéristiques très spécifiques au visage ayant des qualités d'invariance face aux variations des conditions de capture de l'image. Notamment, la pose, l'échelle et l'éclairage. L'hypothèse sous-jacente est basée sur l'observation que les humains peuvent facilement détecter les visages et les objets dans des poses et des conditions d'éclairage différentes. Les caractéristiques souvent utilisées, sont la couleur, la texture et la forme du visage. La couleur de la peau humaine a été utilisée et éprouvé pour être un bon outil de base pour plusieurs systèmes de détection de visage [127] [128] [129] [130] [131] [132]. Bien que différentes personnes aient différentes couleurs de peau, plusieurs études ont

montré que la principale différence réside en grande partie dans leur l'intensité plutôt que leur chrominance. Le principe des approches utilisant la couleur de la peau sera détaillé dans les paragraphes suivants. Ainsi, les visages humains ont une texture distincte qui permet de les séparer de différents objets. Cette propriété est utilisée dans des algorithmes de détection de visages proposés dans [133]. Tandis que, pour la forme du visage, les approches qui exploitent cette caractéristique, procèdent au départ à une détection des contours des objets présents dans l'image. Ces derniers sont ensuite labellisés et mis en correspondance avec un modèle de visage afin de vérifier les détections correctes. De nombreux opérateurs de contours sont utilisés tels que : le filtre de Sobel dans les travaux [134] [135] et le filtre de Canny dans [132].

4.3. Principe de base de l'approche proposée

L'approche que nous proposons pour la détection et localisation de visages a un aspect hybride puisque nous utilisons en premier lieu un détecteur de la couleur de peau (caractéristique invariante) qui couvre toute l'image, et en deuxième lieu, un modèle comme Template pour la classification de visage et non-visage. Le schéma de la figure 4.2 illustre le principe de base de cette approche. Il contient deux modules principaux :

Module 1: la segmentation de l'image couleur pour extraire les régions connexes et significatives représentant des objets d'intérêt composés de pixels de peau permettant ainsi de localiser et extraire les candidats de visage. Ce module se fait en quatre étapes:

- Classifier les pixels de l'image couleur en deux classes (peau/ non peau) en utilisant un détecteur de peau.
- Segmenter l'image binaire résultat de l'étape précédente en régions connexes et cohérentes en utilisant la technique de segmentation connue sous le nom : la Ligne de Partage des Eaux (*LPE, Watershed en Anglais*).
- Elimination des régions non significatives.
- Localisation et extraction des régions de peau significatives

Module 2: la classification de chaque région de peau extraite en tant que candidate de visage : visage ou non-visage, pour décider si un visage est présent ou non dans cette région. Cette tâche est réalisée par la soumission de cette région de peau à une cascade de filtres de Gabor. Ensuite, l'image issue de la sortie de chaque filtre est traitée par un algorithme de

classification de visage ou non-visage en vérifiant sa correspondance avec notre modèle de visage. Et finalement, les régions qui sont classées en tant que visage, sont localisées.

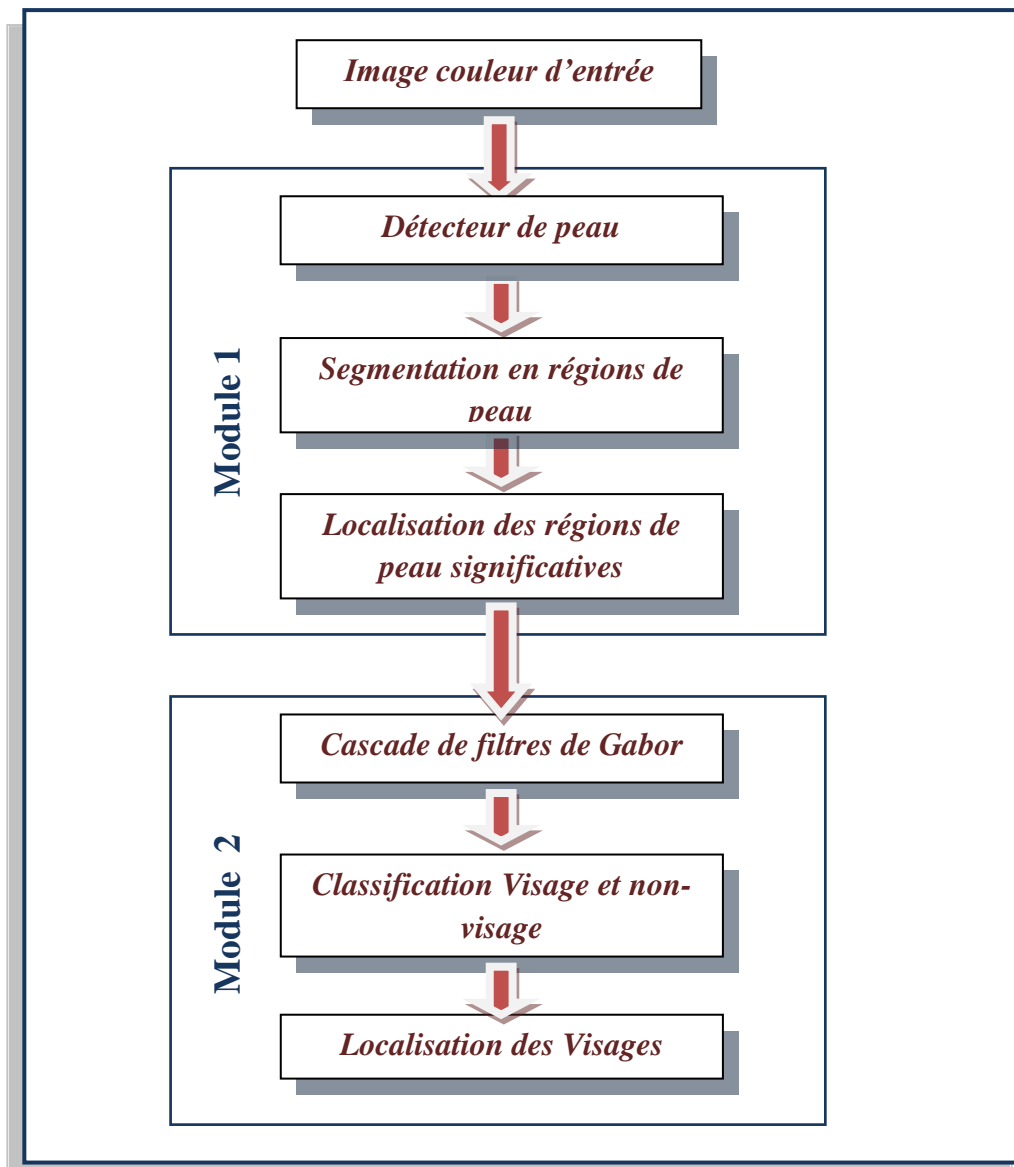


Fig. 4.2. Schéma synoptique de l'approche proposée pour la détection et la localisation de visages

Le choix de concevoir cette approche avec une telle structure est justifié par:

- La robustesse de la couleur de la peau face aux variations géométriques (rotation, changement d'échelle) du modèle de visage [136] et aux occultations partielles.
- L'utilisation de la couleur de la peau comme information de segmentation permet d'une part de réduire l'espace de recherche de la région de visage dans l'image, ce qui a pour effet l'amélioration de la rapidité de détection en comparaison avec d'autres systèmes de détection qui parcourent l'image de manière exhaustive, à toutes les positions et dans toutes les échelles

possibles, telles que les approches basées sur l'apparence. Et d'autre part, de détecter la totalité du visage.

- Le modèle que nous avons proposé comme Template est d'une part, basé sur l'apparence globale et locale du visage et il ne nécessite ni un calcul de corrélation, ni un apprentissage automatique préalable pour assurer la classification visage et non-visage. Et d'autre part, il n'est pas appliqué sur l'image candidate-visage directement, mais sur la réponse de sa convolution avec une cascade de filtres de Gabor. Or, la représentation de visage par des ondelettes de Gabor a montré à plusieurs reprises son aptitude à détecter des caractéristiques particulières dans une image de visage tout en minimisant les effets liés aux variations d'illumination et de pose.

4.4. Détecteur de peau

Le point commun entre les systèmes de détection de visage qui exploitent l'approche de la couleur de la peau est l'utilisation d'un détecteur de peau. Ce dernier est un classificateur qui définit un espace de décision binaire dans l'espace de représentation couleur. Ainsi chaque pixel est classifié en tant que pixel de peau ou pixel non peau (figure 4.3).

La construction d'un détecteur de peau se fait par le choix adéquat d'un espace de couleur et la définition d'un modèle de distribution de la couleur de peau. Ce choix reste toujours difficile puisqu'on vise au départ un espace et un modèle qui couvrent toutes les différentes teintes que peut prendre la couleur de la peau. Cependant, ces teintes peuvent se décliner à d'autres nuances à cause de la variation d'éclairage.

4.4.1. Espaces de couleurs

Pour modéliser la distribution de la couleur de peau, plusieurs espaces de couleurs populaires ont été proposés dans la littérature. Nous citons à titre d'exemple les espaces les plus utilisés et que nous avons comparés expérimentalement pour valider les espaces adoptés dans notre détecteur. Notons que les définitions de ces espaces sont présentées dans l'annexe A.

Le système RGB est largement utilisé dans la détection des pixels de peau dans des images de couleur [130] [132] [137] [138], vu sa simplicité et les résultats satisfaisants qu'il permet d'obtenir.

Beaucoup de chercheurs recommandent aussi le système RGB normalisé [128] [139].

L'espace de couleur HSV a été utilisé pour la détection des pixels de peau dans le travail [129].

Les articles [127] [131] affirment que l'espace de couleur YCbCr permet une bonne détection des pixels de la peau.

4.4.2. Modèles de la distribution de la couleur de la peau

La modélisation de la distribution de la couleur de la peau permet d'attribuer à chaque pixel une vraisemblance selon laquelle il est de couleur peau. Comme l'indiquent les études présentées dans [136] [140], cette distribution est souvent définie avec l'un des trois modèles suivants :

4.4.2.1. Le modèle de peau paramétrique

Ce modèle permet d'ajuster les distributions avec quelques fonctions spécifiques paramétrées telles que: le modèle *uni-gaussien* qui modélise la distribution de la couleur peau par une loi normale dans un espace colorimétrique, le mélange de Gaussiennes qui permet d'obtenir un modèle plus précis, par rapport au modèle uni-gaussien ou un modèle elliptique de borne [141].

4.4.2.2. Le modèle de peau non paramétrique

Ce modèle vise à estimer la distribution de couleurs de peau par des données d'apprentissage sans aucune hypothèse sur la fonction de densité de probabilité de la distribution, tel que le modèle Bayésien basé sur les histogrammes [137].

4.4.2.3. Le modèle de peau à seuils fixes

Ce modèle définit explicitement un certain nombre de règles qui délimitent la zone de la couleur de peau dans un espace de couleurs par plusieurs bornes (seuils fixes) [127] [128] [129] [130] [131] [131] [138]. Quelques exemples de ce modèle sont donnés dans la section suivante.

4.4.3. Le modèle de la couleur de peau proposé

Le choix du modèle de la couleur de peau adapté à une application de segmentation d'une image en couleur est toujours complexe. Dans notre système de détection de visages, la détection de peau est juste une étape préalable et devrait donc être rapide tout en ayant des meilleurs performances en termes de taux de vrais positifs (**VP**: la probabilité qu'un pixel appartenant à la classe peau soit affecté à la classe peau) et taux de faux positifs (**FP**: la probabilité qu'un pixel appartenant à la classe non peau soit affecté à la classe peau).

Pour cette raison, notre choix a été porté sur le modèle de peau à seuils fixes qui produit une classification très rapide vu la simplicité des règles de décision utilisées pour discriminer les pixels de peau de ceux de non peau.

En plus de cette qualité, notre modèle de peau doit faire apparaître les traits caractéristiques du visage (les yeux et la bouche) parce que le modèle (Template) utilisé pour la discrimination du visage et non-visage se base sur l'apparence de ces traits caractéristiques. Ainsi, nous souhaitons pouvoir identifier les pixels de couleur de la peau, malgré les variations de luminosité ambiante, pour cela nous sommes intéressés par la teinte de chaque pixel. C'est à dire l'information de couleur indépendamment de sa luminosité.

Pour aboutir à un modèle de peau avec l'espace de couleur adéquat pour notre système de détection, des tests comparatifs ont été effectués tout d'abord sur quatre modèles de peau existants utilisant différents espaces de couleurs et le modèle de peau à seuils fixes. Ces modèles sont :

- Pour l'espace **RGB**, nous avons testé le modèle proposé par [138] qui prouve que les pixels de la peau doivent vérifier les règles suivantes :

$$\left\{ \begin{array}{l} R > 95 \text{ et } G > 40 \text{ et } B > 20 \\ \max\{R, G, B\} - \min\{R, G, B\} > 15 \\ |R - G| > 15 \\ R > G \text{ et } R > B \end{array} \right. \quad (4.1)$$

- Pour l'espace **RGB normalisé**, nous avons testé le modèle proposé par [128] qui définit le modèle de peau par les bornes suivantes:

$$0.36 \leq r \leq 0.56 \quad \text{et} \quad 0.28 \leq g \leq 0.36 \quad (4.2)$$

- Pour l'espace **YCbCr**, le test est fait avec le modèle de [127] qui fixe les règles suivantes que doit vérifier un pixel de couleur de peau :

$$97.5 \leq Cb \leq 142.5 \quad \text{et} \quad 134 \leq Cr \leq 176 \quad (4.3)$$

- Pour l'espace **HSV**, nous avons testé le modèle proposé par [129] qui définit les plages des valeurs de pixels de peau dans cet espace comme suit :

$$0 \leq H \leq 50, \quad 0.2 \leq S \leq 0.68 \quad \text{et} \quad 0.35 \leq V \leq 1 \quad (4.4)$$

La synthèse de la comparaison des quarts modèles et notre modèle proposé ci-après, avec des exemples de résultats seront exposés dans le chapitre 7. Cette synthèse nous a permis de tirer la conclusion suivante : combiner les deux modèles RGB et RGB normalisé. Et dans le but de préserver les avantages de ces deux modèles, nous avons construit le graphe d'induction (figure 4.3) définissant les règles de cette combinaison.

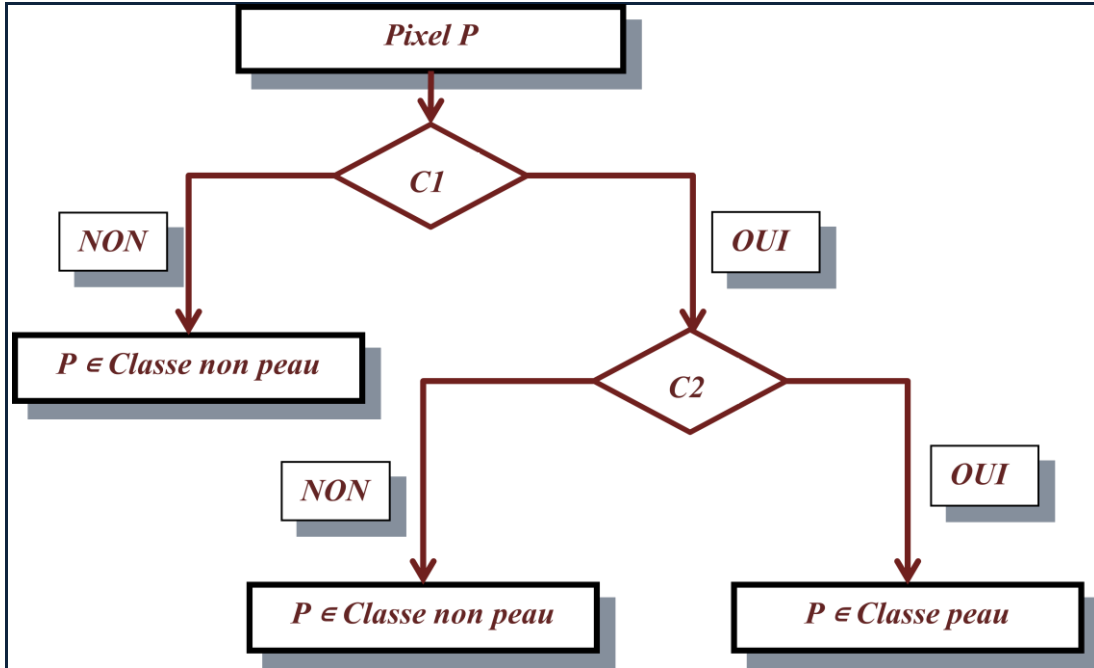


Fig. 4.3. Graphe d'induction du modèle de la couleur de peau proposé

Avec :

<u>Condition C1 :</u>	<u>Condition C2 :</u>
$\left\{ \begin{array}{l} R > 95 \text{ et } G > 40 \text{ et } B > 20 \\ \max\{R, G, B\} - \min\{R, G, B\} > 15 \\ R - G > 15 \\ R > G \text{ et } R > B \end{array} \right. \quad (4.5)$	$\left\{ \begin{array}{l} 0.36 \leq \frac{R}{R+G+B} \leq 0.56 \\ 0.28 \leq \frac{G}{R+G+B} \leq 0.36 \end{array} \right. \quad (4.6)$

Pour finir cette étape de détection des pixels de peau, nous présentons dans la figure 4.4 un exemple de détection des pixels de peau dans une image couleur avec notre modèle de la couleur de peau proposé.

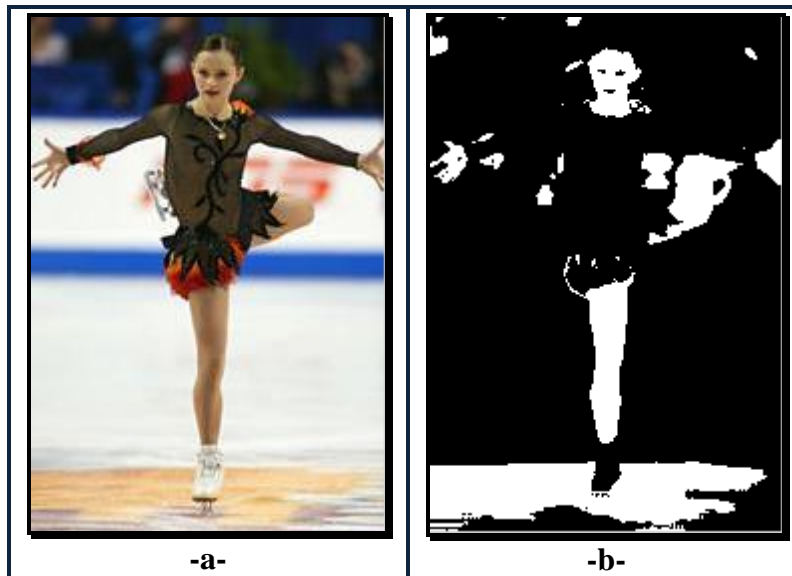


Fig. 4.4. Illustration de l'étape de la détection des pixels de peau : a) image couleur d'entrée et b) image binaire des pixels de peau et non peau.

4.5. Segmentation en régions de peau connexes et significatives

Le détecteur de la peau nous permet de détecter les pixels de peau, ce qui nous donne comme résultat une image binaire constituée des pixels de peau et celles de non-peau. Or, dans notre approche, nous aurons besoin de localiser les régions de peau. Pour cette fin, nous effectuons une segmentation qui consiste à extraire des régions cohérentes et significatives représentant des objets d'intérêt composés de pixels de peau.

La segmentation est un traitement souvent employé pour analyser et interpréter des images. Elle a pour but le partitionnement d'une image en plusieurs régions pertinentes et homogènes. Ainsi, une région correspond à un ensemble de pixels de l'image ayant des propriétés communes (intensité lumineuse, texture, couleur, forme, ...) qui les différencient des pixels des régions voisines.

Nous avons utilisé comme méthode de segmentation de l'image *la ligne de partage des eaux* (LPE). La morphologie mathématique est à l'origine de la définition de cette technique de segmentation.

4.5.1. Présentation de la ligne de partage des eaux (LPE)

La ligne de partage des eaux LPE (Watershed en anglais) est un opérateur de la morphologie mathématique qui est considéré actuellement comme un des outils les plus puissants permettant la segmentation des images en niveaux de gris [142] [143]. La notion de ligne de partage des eaux n'est pas en elle-même à proprement parler une notion purement

issue de la morphologie mathématique, mais trouve son origine en topographie et en hydrologie.

La LPE dans le sens de la topographie est la frontière (la limite géographique) entre deux bassins versants. Cette ligne partage l'écoulement des eaux de surface (sources, précipitations) entre deux directions différentes, chacune correspondant à un bassin versant et un exutoire distincts (figure 4.5). Une goutte d'eau tombant sur ce relief suit un chemin descendant jusqu'à atteindre un minimum régional (Minima). C'est une notion qui apparaît dès le XIX siècle avec notamment, les travaux de J. C. Maxwell. Un siècle plus tard, elle a été introduite par S. Beucher et C. Lantuéjoul [144] pour la segmentation d'image en niveaux de gris. Cette dernière peut être considérée comme un relief topographique : le niveau de gris de chaque pixel dans l'image est interprété comme son altitude dans le relief topographique (figure 4.6). Les bassins et vallées correspondent aux zones sombres de l'image alors que les montagnes et les lignes de crêtes correspondent aux zones claires.

De nombreux auteurs se sont penchés sur cette notion de LPE en essayant de définir des algorithmes permettant son implémentation pour un usage de segmentation d'image. On trouve dans la littérature trois principaux algorithmes à partir desquelles dérivent plusieurs versions: le premier est basé sur l'inondation [144], le deuxième sur l'utilisation des fonctions de distances géodésiques [145] et le troisième sur un algorithme récursif d'immersion [146].

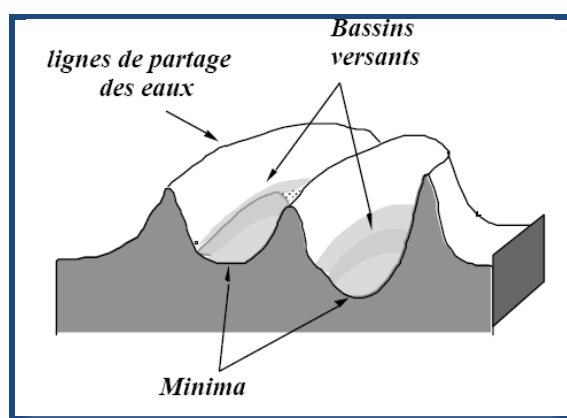


Fig. 4.5. Illustration de la LPE topographique

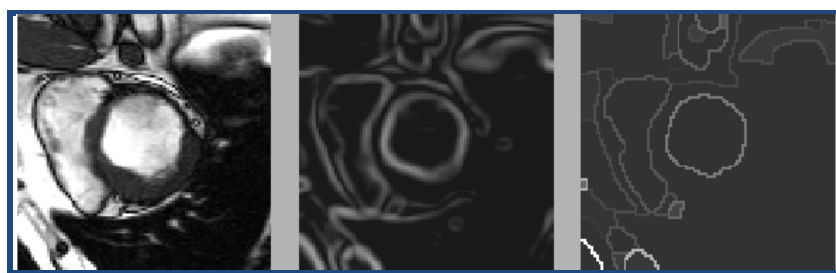


Fig. 4.6. Exemple de segmentation d'une image en niveaux de gris par la LPE

4.5.2. Construction de la ligne de partage des eaux (LPE)

La construction de LPE en segmentation des images, nécessite la connaissance de quelques pré-requis dont les définitions sont données ci-après :

Définition du Chemin géodésique :

Considérons l'ensemble (compact) X dans R^2 de la figure 4.7. Soient deux points P_1 et P_2 appartenant à cet ensemble. Une ligne ou un arc d'extrémités P_1 et P_2 sera appelé chemin géodésique s'il vérifie les conditions suivantes :

- Tous ses points doivent être totalement inclus dans l'ensemble X ;
- La distance entre les deux extrémités P_1 et P_2 (distance géodésique) doit correspondre à la plus petite distance de parcours entre P_1 et P_2 .

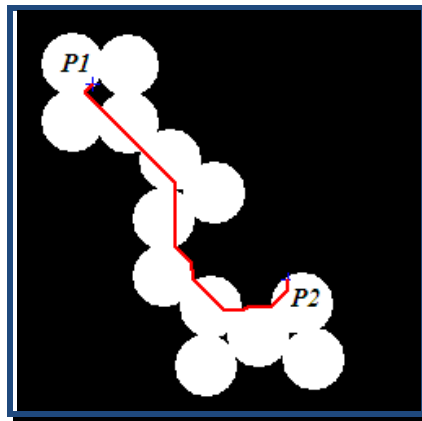


Fig. 4.7. Illustration du chemin géodésique entre les deux points P_1 et P_2 .

Définition de la Zone d'influence géodésique (ZI):

Soit, dans R^2 , l'ensemble $Y = \{Y_i, i \in I\}$ formé de I composantes connexes compactes, toutes incluses dans le compact X (figure 4.8).

La zone d'influence géodésique (ZI) de Y_i dans X se définit comme le lieu des points p de X qui sont géodésiquement plus proches de Y_i que toutes les composantes connexes de X :

$$ZI(Y_i / X) = \{p \in X, \forall k \neq i, D_x(p, Y_i) \leq D_x(p, Y_k)\} \quad (4.7)$$

Où : D_x est la distance géodésique dans l'ensemble X .

Définition du Squelette par zone d'influence (SKIZ):

Le SKIZ est la frontière entre les différentes zones d'influence (figure 4.8) définie par la relation suivante :

$$SKIZ_X (Y_1, \dots, Y_n) = X / \bigcap_1^n ZI_X(Y_i) \quad (4.8)$$

Les zones d'influence et le SKIZ représentent respectivement les bassins versants et la LPE dans le sens de la topographie.

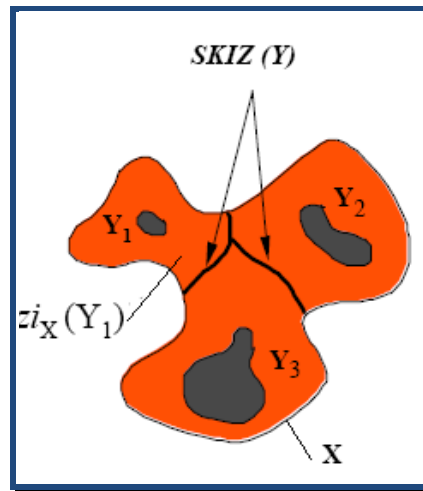


Fig. 4.8. Exemple de SKIZ géodésique

4.5.3. L'algorithme de la LPE adopté

Dans notre approche nous avons opté pour la segmentation des images par la LPE, via l'algorithme récursif d'immersion en raison de sa pertinence et sa simplicité lors de l'implémentation. Cet algorithme est introduit pour la première fois par Vincent et Soille en 1991 [146]. L'algorithme proposé met en application le principe de l'immersion. L'idée est de considérer des seuils successifs de l'image et de calculer des zones d'influence géodésiques d'un niveau à l'autre, le plus rapidement et le plus exactement possible. Pour avoir un accès direct aux pixels d'un niveau de seuil donné, un tri initial des pixels dans l'ordre des niveaux de gris croissants est effectué. Cette première étape exploite le fait que le nombre de niveaux de gris possible dans l'image est fixe : le tri est ainsi effectué en temps linéaire par rapport au nombre de pixels. La deuxième étape, l'inondation, met en jeu des calculs successifs de zones d'influence géodésiques.

4.5.4. Prétraitement avant la segmentation de la LPE

L'algorithme issu de la LPE doit traiter une image en niveaux de gris alors que nous disposons d'une image binaire (peau, non peau). La transformé de distance euclidienne (*EDT, Euclidean Distance Transform*) [147] permet de résoudre ce problème en générant une image en niveaux de gris à partir d'une image binaire et en même temps, elle permet de minimiser l'effet de l'un des problèmes majeurs de la LPE, à savoir : la sur-segmentation. L'algorithme de LPE génère le plus souvent une sur-segmentation importante. La sur-segmentation se produit à cause du grand nombre de minima locaux non significatifs. Il suffit

de petites fluctuations du relief pour produire autant de minima locaux dans l'image. Or, la transformé en distance permet de supprimer les minima à faible contraste, ce qui a pour avantage d'annuler l'effet des bassins non-significatifs. Cette transformé appelée parfois la carte de distances, est une représentation d'une image binaire dans laquelle on a associé à chaque pixel de l'image la distance au point obstacle le plus proche. Autrement dit, étant donnée une distance d et une image binaire contenant une forme X , la transformée de distance est une copie de l'image originale dans laquelle chaque point de X est étiqueté par sa distance à \bar{X} . De nombreux algorithmes de traitement d'images ont recours à cette transformée pour accélérer les calculs tels que les opérateurs morphologiques et les détecteurs de squelette.

Nous présentons dans la figure 4.9 un exemple de segmentation par la LPE d'une image binaire issue du détecteur de peau après sa transformée de distance.

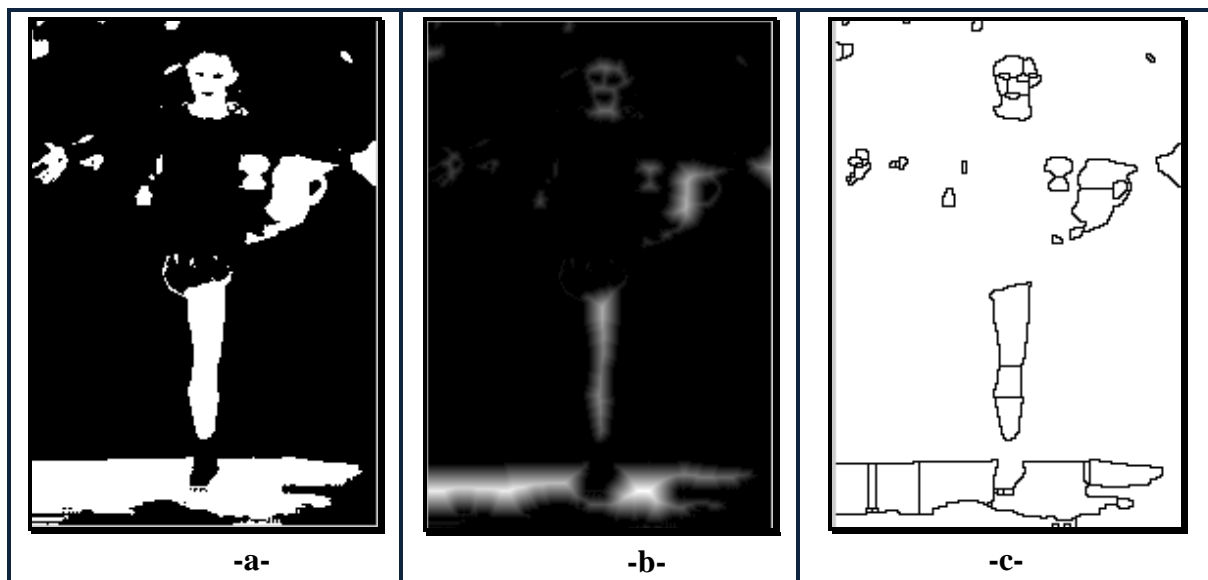


Fig. 4.9. Illustration de la segmentation par la LPE d'une image binaire: (a) image binaire des pixels de peau et non peau, (b) image de transformée de distance et (c) image segmentée par la LPE

4.5.5. Localisation des régions de peau connexes significatives

La segmentation par LPE appliquée à l'image transformée de distance de l'image binaire issue du détecteur de peau étant achevée, nous devons maintenant masquer toutes les régions de peau afin de déterminer les régions connexes les plus significatives qui vont être localisées et éliminer le reste. Pour atteindre cet objectif, il faut tout d'abord rassembler les segments (zones d'influence) adjacents en une seule région connexe. Ensuite, pour localiser ces régions, nous déterminons le nombre de régions de peau dans l'image à partir de leurs masques, en

associant à chaque région une valeur entière appelée étiquette. Une fois, cette étape réalisée, nous extrayons pour chaque région les informations suivantes :

- La position : les coordonnées (X,Y) du coin supérieur gauche de la région par rapport à l'image segmentée.
- Les dimensions : la largeur et la hauteur du segment.

Nous avons choisis comme critère des régions non-significatives qui vont être éliminées, les régions ayant la taille minimale suivante :

$$\text{Largeur} \leq 15 \text{ et Hauteur} \leq 15 \quad (4.9)$$

Ainsi, notre algorithme ne détectera pas un visage de taille inférieure à 15x15. Les autres régions constitueront des visages potentiels. La figure 4.10 illustre les étapes de masque des régions de peau, élimination des régions de peau non-significatives et localisation de celles les plus significatives.

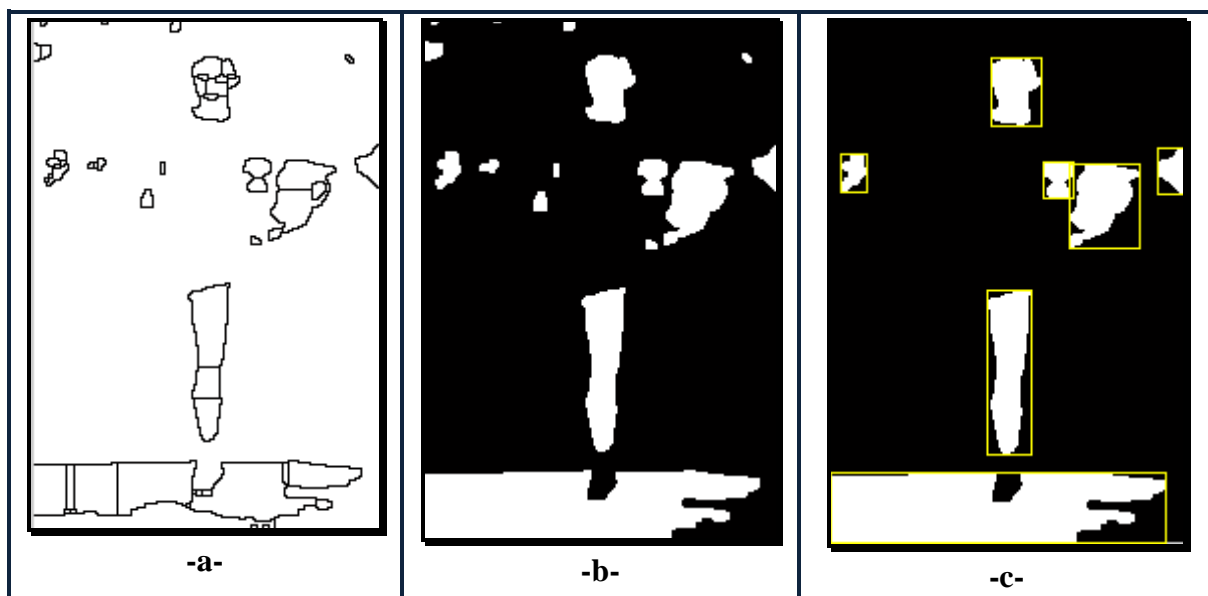


Fig. 4.10. Illustration des étapes de masque et de localisation des régions de peau : (a) image segmentée par la LPE; (b) masque des segments de peau adjacents; (c) image après élimination des régions de peau non significatives et localisation des régions de peau significatives.

4.6. Classification visage et non-visage

Le module de segmentation de l'image couleur en régions de peau connexes permet de localiser et extraire des régions qui contiennent des objets visage ou des objets non-visage tels que : des organes du corps humain ou des objets du fond de l'image qui ont une teinte proche

de la couleur de peau. D'où la nécessité d'intégrer ce module de classification qui permet de discriminer les régions visages de celles non-visages.

Dans ce module chacune des régions de peau localisées dans le module précédent est soumise à une cascade de quatre filtres de Gabor. Ensuite, chaque image issue de la sortie d'un filtre de Gabor est transformée en format binaire pour qu'elle soit traitée par l'algorithme de classification visage et non-visage. A ce moment là, cet algorithme décide si la région de peau traitée est un visage ou non selon le résultat de correspondance avec notre modèle de visage décrit ci-après.

4.6.1. Présentation du Filtre de Gabor (ondelette de Gabor)

Le filtre de Gabor (ou ondelette de Gabor), défini par Dennis Gabor en 1946 qui porte maintenant son nom, est introduit dans le but de palier l'un des défauts de la transformé de Fourier, à savoir l'impossibilité de conserver l'information spatiale lors de la transformation d'un signal. En 1985, John G. Daugman a proposé une extension bidimensionnelle de ce filtre qui l'a utilisé par la suite pour l'analyse des images dans son travail [148].

Le filtre de Gabor fait partie de la catégorie des filtres directionnels qui ont pour objectif d'extraire l'information utile dans la direction recherchée. Pour cela, le filtre de Gabor est très souvent utilisé en traitement d'image car c'est un filtre passe bande simple à concevoir et possède une résolution conjointe spatiale/fréquentielle optimale, c.à.d. qu'il offre la meilleure localisation simultanée en temps et en fréquence.

Mathématiquement, Le filtre 2-D de Gabor est une gaussienne bidimensionnelle modulée par une fonction sinusoïdale plane. L'expression d'un tel filtre dans le domaine spatial (x, y) est définie par l'expression:

$$G(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\left[\left(\frac{1}{\sigma_x^2}\right)x'^2 + \left(\frac{1}{\sigma_y^2}\right)y'^2\right]} e^{j2\pi f_0 x'} \quad (4.10)$$

$$\text{Avec : } \begin{cases} x' = x \cdot \cos \theta + y \cdot \sin \theta \\ y' = -x \cdot \sin \theta + y \cdot \cos \theta \end{cases} \quad (4.11)$$

Où σ_x et σ_y sont les écart-types de la gaussienne respectivement le long des axes x et y , f_0 est la fréquence centrale du filtre et θ l'orientation de l'ondelette (angle entre l'axe des abscisses et le vecteur de module f_0).

La figure 4.11 illustre la partie réelle d'un exemple de filtre de Gabor ayant les paramètres: $\sigma_x = \sigma_y = 3$, $f_0 = 0.25$ et $\theta = \pi/6$.

Le filtre de Gabor possède quatre degrés de liberté :

- L'orientation du filtre θ : ce paramètre fait pivoter l'ondelette autour de son centre. L'orientation de l'ondelette (Figure 4.12) détermine l'angle des contours ou des lignes de l'image auxquels l'ondelette va être sensible.

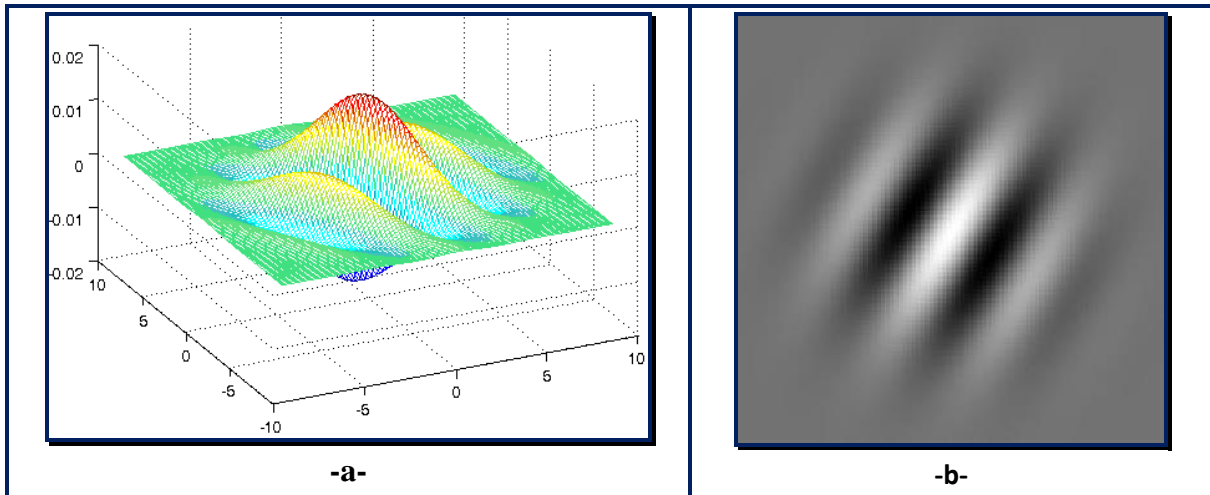


Fig. 4.11. La partie réelle d'un exemple de filtre de Gabor : a) Vue en 3D et b) vue en 2D de dessus (les axes (x,y) représente le plan spatial et l'axe z l'amplitude du filtre)

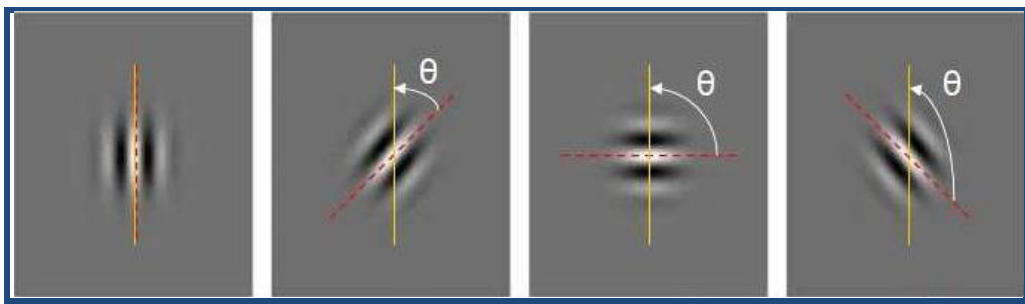


Fig. 4.12. La représentation de la partie réelle de 4 filtres de Gabor ayant respectivement l'orientation θ : $0; \pi/4; \pi/2; 3\pi/4$.

- La fréquence centrale f_0 que l'on cherche à extraire de l'image: Le choix de la fréquence centrale f_0 du filtre est très important car la qualité de l'image finale dépend directement du bon choix de ce paramètre. Les ondelettes avec une petite fréquence centrale seront sensibles à des changements progressifs d'intensité dans une image. Les ondelettes avec une grande fréquence centrale seront sensibles à des contours et des lignes abruptes. La figure 4.13 illustre la partie réelle de quatre filtres avec différentes valeurs de la fréquence centrale f_0 .

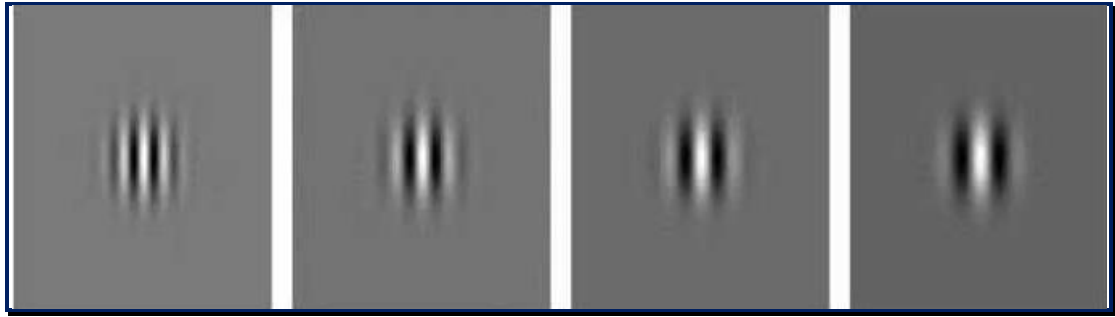


Fig. 4.13. La représentation de la partie réelle de 4 filtres de Gabor ayant respectivement la fréquence centrale f_0 : $1/8$; $1/12$; $1/14$; $1/16$.

- Les écarts types permettent de régler les bandes passantes fréquentielle et directionnelle du filtre (figure 4.14), c.à.d les rayons de la gaussienne. Or, La taille de la gaussienne détermine la quantité de pixels de l'image à filtrer qui seront pris en compte dans la convolution.

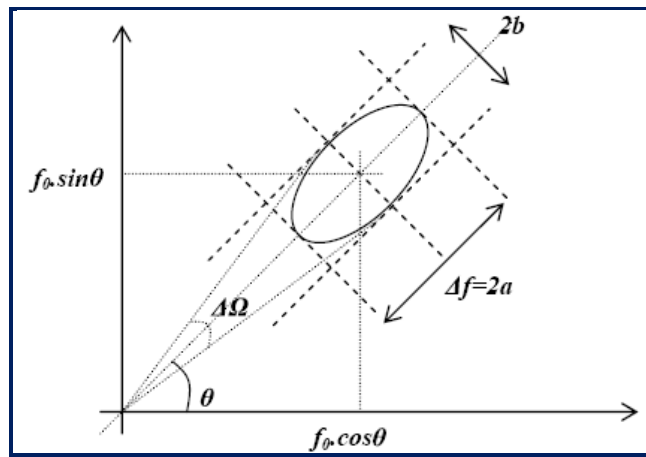


Fig. 4.14. Représentation des bandes passantes fréquentielles (Δf) et directionnelles ($\Delta\Omega$) d'un filtre de Gabor (a et b sont les rayons de l'ellipse centrée en $(f_0 \cos \theta, f_0 \sin \theta)$).

La représentation en ondelettes de Gabor d'une image I (Filtrage de cette image par un filtre de Gabor) résulte du produit de convolution de l'image I avec ce filtre comme il est défini par l'équation suivante :

$$F(x, y) = I(x, y) * G(x, y) \quad (4.12)$$

4.6.2. Filtres de Gabor en modélisation faciale

Les filtres 2-D de Gabor ont été introduits dans le domaine de la reconnaissance de visage pour la première fois par Lades et al [84] pour la détection des caractéristiques de visage. Ensuite, ils sont devenus un outil puissant dans plusieurs algorithmes de reconnaissance de visage comme nous l'avons cité dans l'état de l'art et de détection de visage [115] [116]. L'extraction de l'information par les filtres de Gabor a été montrée dans toutes ces

publications comme très pertinente au vu des performances obtenues. En effet, ces travaux affirment que la représentation par ondelettes de Gabor des images faciales est robuste aux changements causés par des variations d'éclairage ou par des modifications d'expressions faciales. Ainsi, elle est capable de minimiser l'effet de la variation de pose dans une marge qui ne dépasse pas 40°.

La majorité des ces travaux utilisent un banc des filtres de Gabor construits à partir de l'expression paramétrée suivante:

$$G_{u,v}(x, y) = \frac{f_u^2}{\pi k \rho} e^{-\left[\left(\frac{f_u^2}{k^2}\right)x'^2 + \left(\frac{f_u^2}{\rho^2}\right)y'^2\right]} e^{j2\pi f_u x'} \quad (4.13)$$

$$\text{Avec : } f_u = f_{max}/2^{\left(\frac{u}{2}\right)} \text{ et } \theta_v = v\pi/8 \quad (4.14)$$

f_u et θ_v désignent respectivement la fréquence centrale et l'orientation de l'onde plane paramétrées par u et v . Les paramètres k et ρ déterminent le rapport entre la fréquence centrale et la taille de l'enveloppe gaussienne le long des axes x et y respectivement (c.à.d les écart-types de la gaussienne σ_x et σ_y). Le choix de différentes valeurs pour ces paramètres permet de construire un banc de filtres de Gabor.

4.6.3. Filtres de Gabor dans notre approche de détection de visage

Dans notre approche de détection de visage, nous utilisons les filtres de Gabor dans le deuxième module pour la classification des régions de peaux localisées en deux classes visage et non-visage à l'aide d'une cascade de filtres. Cette dernière vise la vérification de la présence et la localisation des traits caractéristiques du visage (les yeux et la bouche) dans une région de peau. Pour choisir les filtres de la cascade qui seront optimaux pour réaliser cette tâche, nous avons tout d'abord construit un banc de 40 filtres à partir de l'expression 4.13, correspondant à :

- 8 orientations: $v = \{0, 1, 2, \dots, 7\}$
- 5 fréquences centrales : $u = \{0, 1, 2, \dots, 5\}$ avec: $f_{max} = 0.25$

Ensuite, nous avons effectué une série d'opérations de filtrage à l'aide de ce banc sur 80 images de visage. Ces tests nous ont permis de conclure que le filtre qui détecte mieux la présence et l'emplacement des traits caractéristiques du visage doit avoir un angle d'orientation θ_v qui vérifie la relation suivante :

$$\theta_v \simeq \frac{\pi}{2} + \lambda \quad (4.15)$$

Avec : λ est l'angle d'inclinaison de la région de la peau par rapport à l'axe verticale passant par son centre de gravité G (figure 4.15).

En plus, la fréquence centrale de ce filtre dépend d'une part de la taille du visage et d'autre part de la qualité de l'image de sortie du détecteur de peau. Les résultats présentés dans les figures 4.15 et 4.16 affirment cette conclusion.

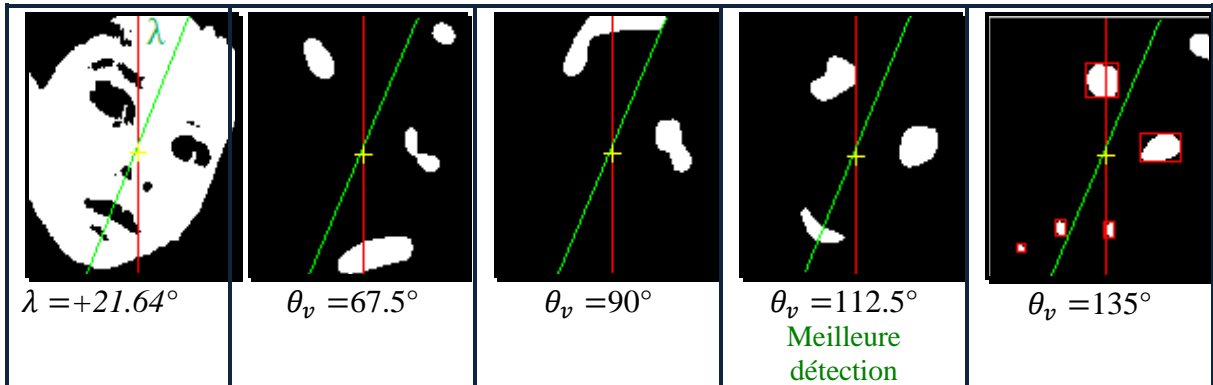


Fig. 4.15. Résultat de filtrage d'une image de visage qui a un angle d'inclinaison $+21.64^\circ$ par des filtres ayant la même fréquence centrale et différentes valeurs d'orientation θ_v

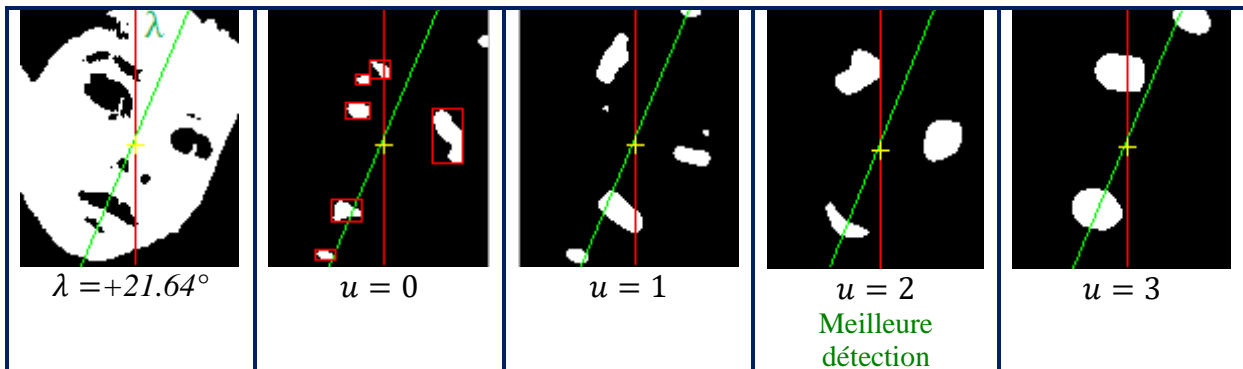


Fig. 4.16. Résultat de filtrage d'une image de visage avec un angle d'inclinaison $+21.64^\circ$ par des filtres ayant la même valeur de θ_v et différentes valeurs de la fréquence centrale f_u

En partant de cette conclusion, on doit effectuer les deux actions suivantes:

- Calculer l'angle d'inclinaison λ de la région de peau localisée.
- Appliquer une rotation d'un angle de $-\lambda$ à la région de peau.
- Soumettre l'image binaire, sortie du détecteur de la peau, à une cascade de 4 filtres ayant respectivement les fréquences centrales : $u = 0$; $u = 1$; $u = 2$; $u = 3$ (la fréquence centrale du filtre est paramétrée par u) et ayant la même orientation $\theta_v = 90^\circ$ (c.à.d $v = 4$).

Pour déterminer les coordonnées du centre de gravité G et l'angle d'inclinaison λ de la région de peau localisée en question, nous utilisons les formules suivantes :

Centre de gravité $G(x_G, y_G)$:

$$x_G = \frac{1}{A} \sum_{i=1}^n \sum_{j=1}^m j \cdot R[i, j] \quad (4.16)$$

$$y_G = \frac{1}{A} \sum_{i=1}^n \sum_{j=1}^m i \cdot R[i, j] \quad (4.17)$$

Où R est la matrice $[n, m]$ du rectangle de l'image contenant la région de peau et A est l'air de cette région (le nombre de pixels appartenant à cette région).

Angle d'inclinaison λ :

Habituellement les visages sont orientés verticalement, cependant on peut avoir une inclinaison d'un angle λ , ce qui nous ramène à calculer cette angle pour rendre droite la région de peau localisée (ou visage), par la formule suivante :

$$\lambda = \frac{1}{2} \arctan \frac{b}{a-c} \quad (4.18)$$

Où :

$$a = \sum_{i=1}^n \sum_{j=1}^m (x'_{ij})^2 \cdot R[i, j] \quad (4.19)$$

$$b = 2 \sum_{i=1}^n \sum_{j=1}^m x'_{ij} \cdot y'_{ij} \cdot R[i, j] \quad (4.20)$$

$$c = \sum_{i=1}^n \sum_{j=1}^m (y'_{ij})^2 \cdot R[i, j] \quad (4.21)$$

et:

$$x'_{ij} = x_{ij} - x_G \quad (4.22)$$

$$y'_{ij} = y_{ij} - y_G \quad (4.23)$$

4.6.4. Modèle de visage

L'étape finale de ce module est la décision sur la classe d'appartenance de la région de peau localisée (visage ou non-visage). Pour atteindre cet objectif, nous avons élaboré un modèle de visage auquel doit correspondre l'image binaire de la sortie d'un filtre de Gabor de la cascade pour l'affecter à la classe visage. Ce modèle illustré dans la figure 4.17 exige la vérification simultanée des conditions suivantes :

- **Condition C1:** L'image binaire de la sortie du filtre doit contenir au moins trois trous significatifs formant un triangle.
- **Condition C2:** L'axe vertical passant par le centre de gravité du masque de la région de peau doit traverser le trou le plus bas des trois trous.

- **Condition C3** : Les deux autres trous restants doivent appartenir respectivement aux deux quadrants supérieurs Q1 et Q2 formés par les deux axes perpendiculaires passant par le centre de gravité G du triangle.
- **Condition C4** : Les deux distances D_1 et D_2 de ces deux derniers trous par rapport au premier trou doivent vérifier la contrainte suivante :

$$D_1 - D_2 < 0.25 \times \text{Max}(D_1, D_2) \quad (4.23)$$

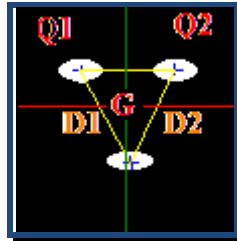


Fig. 4.17. Modèle de visage

4.6.5. Organigramme de décision visage ou non visage

Le module de segmentation permet de segmenter l'image couleur en N régions de peau connexes et significatives. Alors, pour chaque région de peau RP_i (avec: $i=1..N$) localisée et extraite, on effectue le traitement décrit par l'organigramme de la figure 4.18 afin de décider si cette région de peau contient un visage. Si c'est le cas, nous localisons la zone exacte du visage.

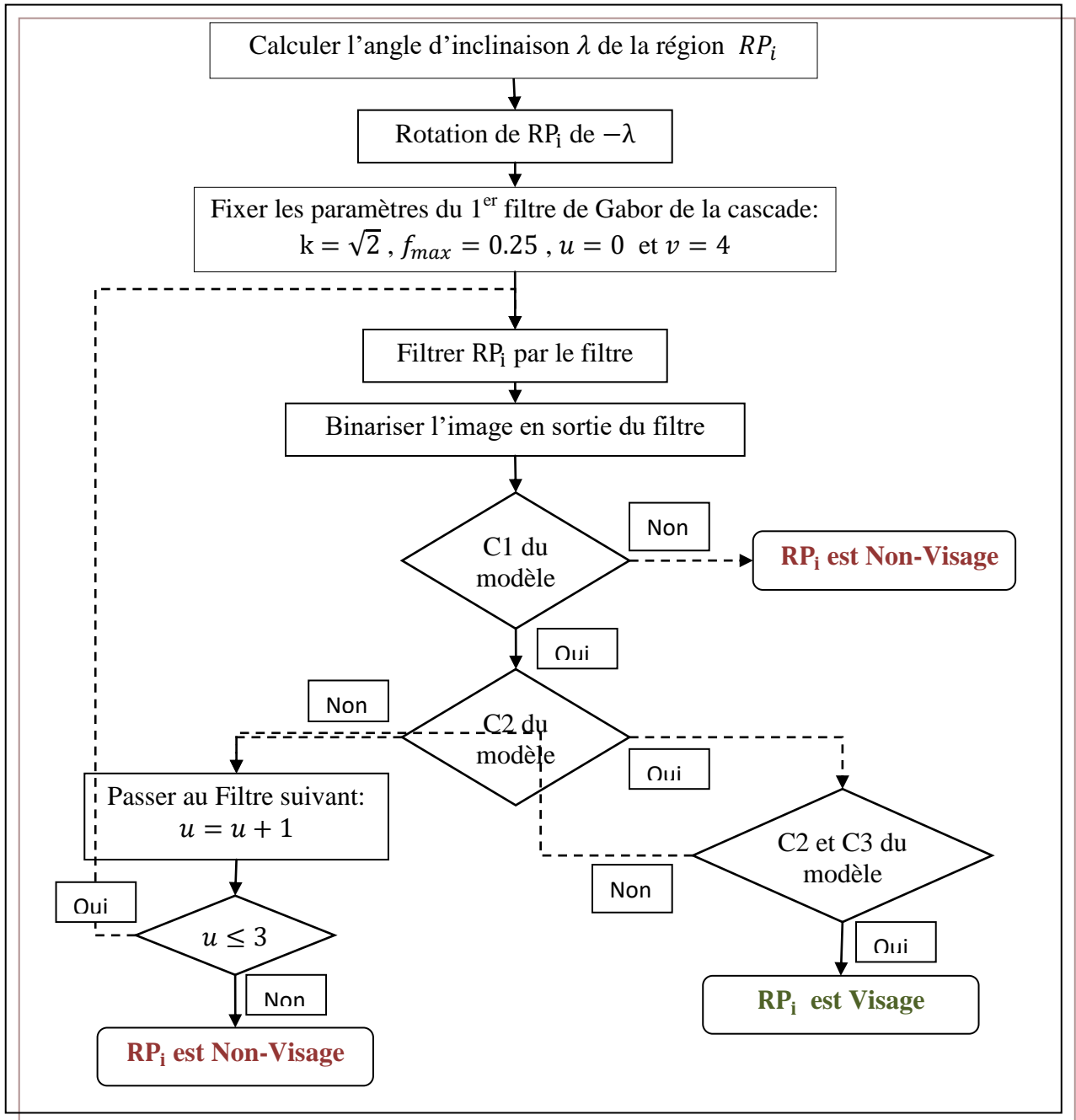


Fig. 4.18. Organigramme de décision visage ou non visage

4.7. Évaluation de la détection et la localisation des visages

Pour évaluer un algorithme de détection de visages, les trois critères les plus importants à prendre en considération sont :

- Le Taux de Vraies Acceptations (**RAR**, **Right Acceptance Rate**) qui exprime le pourcentage des vrais visages acceptés par l'algorithme.

- Le Taux de Fausses Acceptations (**FAR**, *False Acceptance Rate*) qui exprime le pourcentage des régions non-visage acceptés par l'algorithme alors qu'ils devraient être rejetés.
- Le temps de détection de visage.

De plus, un visage est considéré correctement détecté et localisé, si la taille d'image extraite ne dépasse pas de 20% la taille réelle de la région faciale, et qu'elle contient essentiellement les yeux, le nez et la bouche.

4.8. Normalisation

L'étape de normalisation permet de ramener toutes les images de visages détectés à un format prédéfini de telle sorte qu'elles soient exploitables de la même façon dans l'étape d'extraction des caractéristiques (signatures). Il est important que toutes ces images soient de taille identique, au même angle d'inclinaison de visage (généralement vertical) et au même format concernant les couleurs (généralement les images sont converties en niveau de gris). Et enfin pour chaque image, nous essayons de réduire les effets d'illumination. Ce prétraitement peut être divisé en deux processus : normalisation géométrique et normalisation photométrique.

4.8.1. Normalisation géométrique

Cette normalisation est nécessaire. Elle consiste généralement en :

- Un redimensionnement de la taille de l'image de visage à une taille fixe prédéfini parce que la taille du visage à l'intérieur de l'image acquise peut varier en fonction de la distance entre le capteur d'image et la personne.
- Une rotation du visage en position verticale pour annuler l'angle d'inclinaison si celle-ci est différente de 0° , ce qui revient à aligner l'axe des yeux avec l'axe horizontal. Cette opération dans notre système est déjà faite dans l'étape de détection de visage.
- Une élimination des zones non informatives de l'arrière plan et centrage du visage. Cette opération est, de même effectuée dans l'étape de détection de visage. En effet, notre approche de détection se base en premier lieu sur la couleur de peau. Donc, nous détectons et nous localisons uniquement le visage sans arrière plan.

4.8.2. Normalisation photométrique

Dans ce deuxième type de normalisation nous tentons d'éliminer ou réduire les effets de variation d'illumination dans l'image de visage. Pour atteindre cet objectif, nous avons opté pour la méthode de l'égalisation d'histogramme.

L'égalisation d'histogramme a pour but d'harmoniser la répartition des niveaux de gris d'une image. Elle permet de tendre vers un même nombre de pixels pour chacun des niveaux de gris de l'histogramme. Ce traitement vise à augmenter les nuances dans l'image et donc son contraste. Dans la figure 4.19 ci-dessous, nous présentons un exemple d'opération d'égalisation d'histogramme.

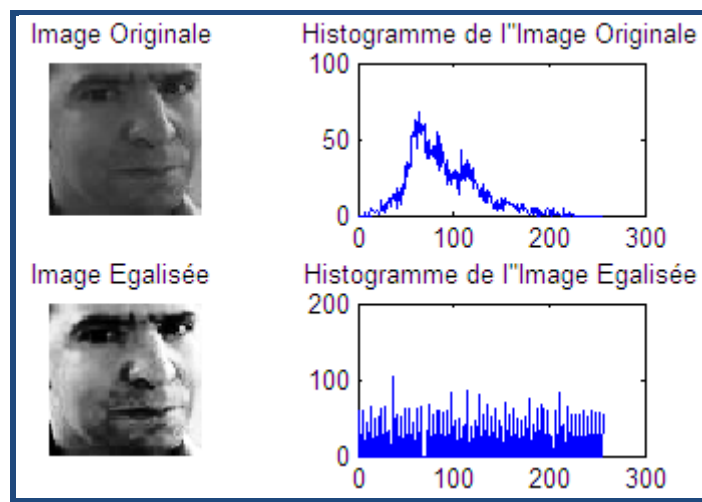


Fig. 4.19. Exemple d'opération d'égalisation d'histogramme : en haut l'image en niveau de gris avant égalisation et en bas après égalisation

4.9. Conclusion

Dans ce chapitre, nous avons présenté une solution robuste de la détection et la localisation de visage basée d'une part sur la segmentation de l'image couleur en régions de peau à l'aide d'une nouvelle approche de détection des pixels de la peau et la méthode de segmentation LPE. Et d'autre part, sur la classification des candidats de la segmentation en deux classes visage et non-visage au moyen des filtres de Gabor combinées avec un modèle de visage. Tous ces différents outils sont bien intégrés dans notre approche pour développer un détecteur de visage ayant des performances très raisonnables que nous allons présenter et discuter dans le chapitre 7. Dans le chapitre suivant, nous nous intéresserons à la description des approches d'extraction des caractéristiques proposées.

Chapitre 5

Extraction des caractéristiques

5.1. Introduction

Après avoir présenté dans le chapitre précédent notre approche de détection de visage, nous présentons dans le présent chapitre nos deux contributions pour l'extraction des caractéristiques. Cette dernière, comme nous l'avons signalé dans le chapitre 2, est une étape particulièrement importante dans tout système de reconnaissance. Elle vise à extraire de l'image de visage détecté, un vecteur caractéristique (signature) qui soit suffisamment représentative d'un visage donné et qui le modélise de façon beaucoup plus précise que l'image brute de départ. Cette nouvelle représentation du visage doit avoir en même temps la propriété d'unicité pour chaque personne et la propriété de discrimination entre plusieurs personnes différentes (pas de risque de confusion entre deux personnes différentes). De plus, cette représentation doit être invariante devant les modifications qui peuvent survenir sur l'apparence de visage au moment de la capture de l'image.

A l'heure actuelle, un tel algorithme idéal est loin d'être réalisé car il est très difficile de développer un algorithme qui soit général tolérant toutes les variations possibles de l'apparence de visage. C'est pourquoi, la majorité des approches proposées font des hypothèses sur l'environnement dans lequel peut être appliqué chacune de ces approches.

En cherchant les deux capacités de discrimination et de robustesse qui ont une influence notable sur la performance de fiabilité du système de reconnaissance, il ne faut pas négliger le temps de calcul mis par l'algorithme, surtout en phase de reconnaissance. Mais, il faut réaliser un compromis entre les deux (fiabilité et rapidité).

Pour décrire les approches proposées dans ce sens, nous prenons en considération les notations suivantes:

- On dispose d'une base de données Ω contenant M images de visage, avec plusieurs vues par personne.
- La base de données contient C classes (C personnes).
- L'ensemble des images qui correspondent à une même personne compose une classe Ω_c et chaque classe contient n_c images
- Chaque image de visage est de taille $n \times m$.

5.2. Première approche d'extraction: ACP-ADL double

5.2.1. Principe de base de l'approche

Cette approche fait partie de la catégorie des méthodes globales car elle agit sur la globalité de l'image. Et c'est une technique de réduction de dimension qui cherche un sous-espace de projection de faible dimension qui représente au mieux les variations du visage. Un visage à reconnaître sera projeté sur ce sous-espace pour simplifier la classification. Dans cette approche *ACP-ADL double*, nous avons combiné l'ACP (Analyse en Composantes Principales ou "*EigenFaces*") et de l'ADL (Analyse Discriminante Linéaire) doublée deux fois successives. Le but de cette combinaison est de préserver les avantages des deux types de méthodes. D'une part, ces deux algorithmes s'appuient sur des propriétés statistiques bien connues qui utilisent l'algèbre linéaire. Ils sont relativement rapides à mettre en œuvre surtout en phase de reconnaissance (phase en ligne) qui consiste en une simple projection (produit matriciel). D'autre part, L'ACP permet de réduire la dimensionnalité de l'espace de représentation de l'image de visage tout en optimisant la conservation des informations pertinentes contenues dans l'image et l'ADL permet de trouver des caractéristiques qui séparent relativement bien plusieurs classes.

5.2.2. Description détaillée de l'approche

Dans cette approche, nous définissons trois nouveaux sous-espaces dans lesquelles nous procédons à trois projections successives des images de visage. La figure 5.1 décrit en gros les étapes à suivre dans la phase d'enrôlement pour calculer les matrices de projection modélisant les sous-espaces de projection. Ensuite, elle décrit le processus de projection de l'image de visage à reconnaître permettant d'obtenir finalement son vecteur caractéristique. Ces différentes étapes sont présentées en détail dans les deux sections suivantes.

5.2.2.1. Phase d'enrôlement

Cette phase vise l'extraction des vecteurs caractéristiques des images de visage détectées et normalisées des personnes connues par le système (Base de données ou Galerie). Ce sont ces vecteurs qui seront appris par la suite pour être utilisées plus tard dans la phase de classification. En même temps, dans cette phase, nous déterminons les matrices de projection qui seront utilisées pour le calcul du vecteur caractéristique du visage de la personne à reconnaître.

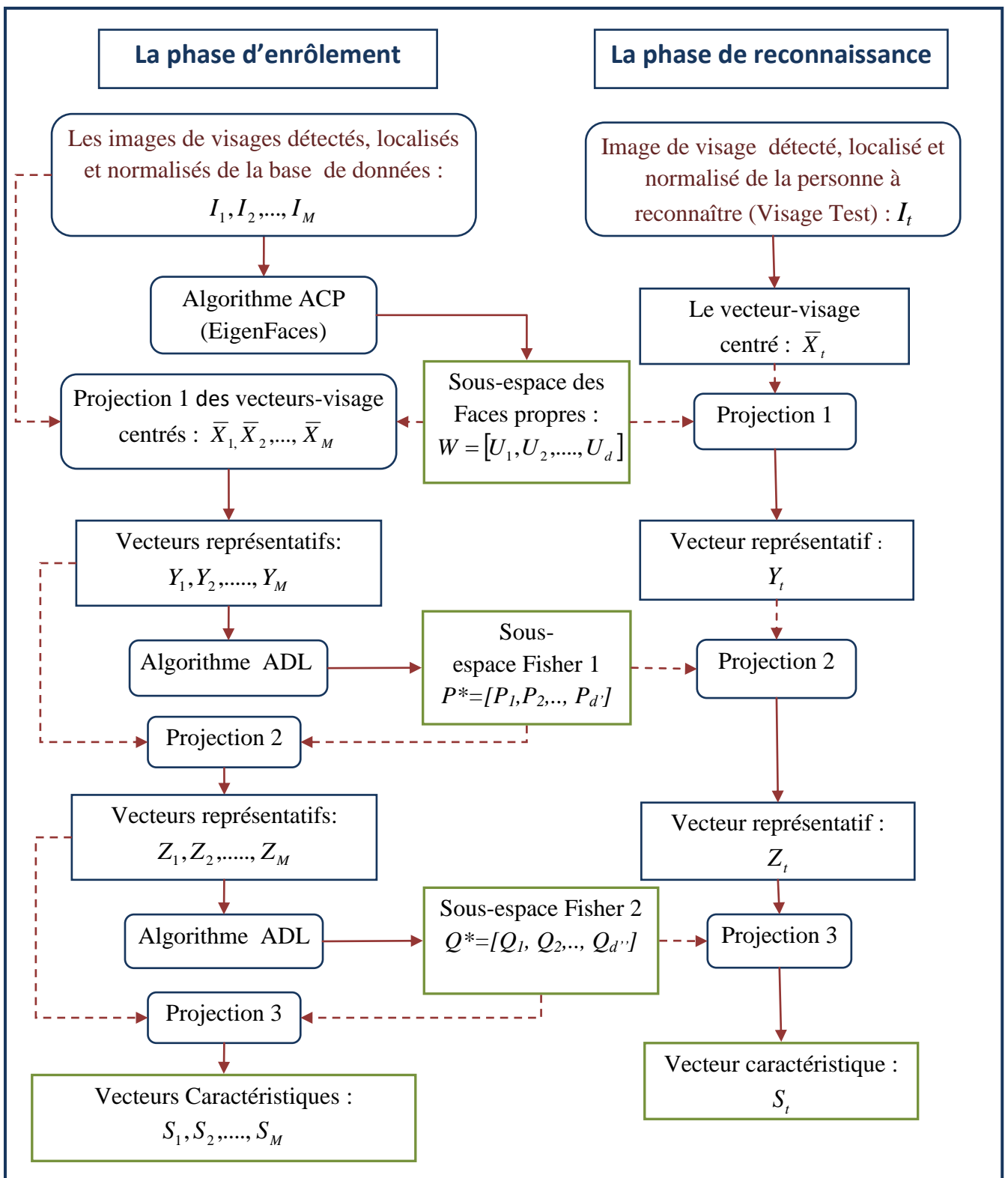


Fig. 5.1. Schéma synoptique de la première approche d'extraction des caractéristiques

Sous-espace de projection 1 (Sous-espace des faces propres)

La méthode ACP (*EigenFaces*) permet d'obtenir un nouvel sous-espace de projection engendré par les faces propres (des vecteurs contenant des informations indépendantes d'un vecteur à l'autre), dans lequel une image peut être réduite à un vecteur de dimension très inférieure tout en garantissant une reconstruction optimale dans le sens inverse. Pour déterminer l'espace des faces propres, on réalise le traitement suivant :

A partir des images de visage de la base de données I_1, I_2, \dots, I_M , on construit la matrice de données de dimension $(N, M) : T = (X_1, X_2, \dots, X_i, \dots, X_M)$. Où chaque X_i est un vecteur colonne de dimension N ($N = n \times m$) représentant l'image de visage I_i après concaténation de ses n lignes ou ses m colonnes.

On calcule ensuite le vecteur moyen (l'image moyenne) μ de l'ensemble des vecteurs X_i . Cette image peut être vue comme le centre de gravité du jeu d'images :

$$\mu = \frac{1}{M} \sum_{i=1}^M X_i \quad (5.1)$$

On ajuste ensuite les vecteurs X_i par rapport à la moyenne. Le vecteur moyen est alors soustrait de chaque vecteur-visage pour déterminer le vecteur-visage centré donné par l'expression suivante:

$$\bar{X}_i = X_i - \mu \quad (5.2)$$

On note par la suite : $A = (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_i, \dots, \bar{X}_M)$ la matrice des vecteurs-visages centrés.

On calcule ensuite la matrice de covariance des données centrées. Cette matrice peut être vue comme une matrice de moments d'ordre 2 :

$$S_T = A^T . A \quad (5.3)$$

L'étape suivante, consiste à calculer les vecteurs propres v_i ($i = 1, \dots, M$) et les valeurs propres λ_i ($i = 1, \dots, M$) de cette matrice de covariance S_T .

Les vecteurs propres trouvés sont ensuite ordonnés selon leurs valeurs propres correspondantes, de manière décroissante. Plus une valeur propre est grande, plus la variance capturée par le vecteur propre associé est importante. Cela implique que la majeure partie des informations est contenue dans les premiers vecteurs propres.

Une part de la grande efficacité de l'algorithme ACP vient de l'étape suivante qui consiste à ne sélectionner que les d meilleurs vecteurs propres (ceux avec les d plus grandes valeurs propres selon un certain critère ou seuil).

A partir des vecteurs propres, on définit les M visages propres U_i qui sont déterminés par l'expression:

$$U_i = \sum_{k=1}^M \bar{X}_k \cdot v_{ik} \quad (5.4)$$

Avec : v_{ik} désigne la $k^{ième}$ composante du $i^{ième}$ vecteur propre.

Et après normalisation des faces propres U_i :

$$U_i^T \cdot U_j = 0, i \neq j, i, j = 1, \dots, M \quad (5.5)$$

On choisit les d faces propres correspondantes aux d valeurs propres dominantes λ_i , pour construire l'espace des visages propres ("*Eigenfaces*" en anglais) défini par la matrice de projection :

$$W = [U_1, U_2, \dots, U_d] \quad (5.6)$$

La dimension d du sous-espace est déterminée de telle façon à réduire autant que possible la dimension du nouvel espace et sans perdre trop d'information. La solution généralement adoptée est de sélectionner le nombre de vecteurs tel que la fraction de la variance totale représente un pourcentage donné d'information. Cette fraction est donné par :

$$q_d = \left(\sum_{i=1}^d \lambda_i \right) / \left(\sum_{i=1}^M \lambda_i \right) \quad (5.7)$$

Finalement, chaque vecteur-visage centré \bar{X}_i de dimension N est réduit à un vecteur Y_i dans le nouvel espace de dimension d en faisant la projection suivante :

$$Y_i = [y_1, y_2, \dots, y_d]^T = W^T \cdot \bar{X}_i \quad i = 1, \dots, M \quad (5.8)$$

Où le vecteur Y_i désigne le vecteur représentatif modélisant l'image-visage I_i .

Sous-espace de projection 2 (Sous-espace Fisher 1):

Maintenant on détermine un deuxième sous-espace de projection (l'espace Fisher 1). Pour le faire, on applique l'algorithme de l'analyse discriminante linéaire de Fisher (ADL) sur l'ensemble des vecteurs Y_i ($i = 1, \dots, M$) obtenus après la projection 1. Cet algorithme calcule un espace de projection (Espace Fisher) qui maximise la distance entre des classes différentes

tout en minimisant la distance entre les éléments provenant d'une même classe. Il s'agit donc de calculer la matrice P^* qui maximise le critère de Fisher généralisé suivant:

$$P^* = \underset{P \in \mathbb{R}^{m \times d^3}}{\text{Arg max}} \cdot \frac{|P^T S_b P|}{|P^T S_w P|} \quad (5.9)$$

Où S_w et S_b désignent respectivement les matrices de covariance intra et inter classes généralisées de la base de données contenant les vecteurs Y_1, Y_2, \dots, Y_M :

$$S_w = \sum_{c=1}^C \sum_{Y_i \in \Omega_c} (Y_i - \bar{Y}_c) \cdot (Y_i - \bar{Y}_c)^T \quad (5.10)$$

$$S_b = \sum_{c=1}^C n_c (\bar{Y}_c - \bar{Y}) \cdot (\bar{Y}_c - \bar{Y})^T \quad (5.11)$$

\bar{Y}_c désigne le vecteur moyen des n_c vecteurs Y_i appartenant à la classe Ω_c et \bar{Y} le vecteur moyen de tous les vecteurs Y_i de la base de données Ω .

Sous l'hypothèse que S_w soit inversible (on peut aisément montrer que cette hypothèse est généralement vérifiée), les d' colonnes de la matrice P^* sont les d' premiers vecteurs propres de la matrice $S_w^{-1} \cdot S_b$ (C'est-à-dire ceux associés aux plus grandes valeurs propres).

Après détermination de l'espace Fisher 1 défini par la matrice de projection $P^*=[P_1, P_2, \dots, P_{d'}]$, on applique la projection linéaire des vecteurs Y_i de taille d en utilisant la formule suivante :

$$Z_i = [z_1, z_2, \dots, z_{d'}]^T = P^{*T} \cdot Y_i \quad i = 1, \dots, M \quad (5.12)$$

Où le vecteur Z_i de taille d' désigne le vecteur signature associé au vecteur Y_i (par transitivité, associé à l'image-visage I_i).

Sous-espace de projection 3 (l'espace Fisher 2):

Cette troisième projection avec le même algorithme ADL est faite dans le but d'améliorer de plus la discrimination entre les données. Cet effet sera justifié par les résultats présentés dans le chapitre 7.

En répétant les mêmes opérations pour trouver le sous-espace de projection 2 (l'espace Fisher 1), mais cette fois-ci avec les vecteurs Z_i ($i = 1, \dots, M$) résultat de la projection 2, on obtient le troisième sous-espace de projection (l'espace Fisher 2) défini par la matrice de projection $Q^*=[Q_1, Q_2, \dots, Q_{d'}]$.

La projection des vecteurs Z_i dans ce troisième sous-espace s'obtient en appliquant l'expression :

$$S_i = [s_1, s_2, \dots, s_{d''}]^T = Q^{*T} \cdot Z_i \quad i = 1, \dots, M \quad (5.13)$$

Ceci donne finalement les vecteurs caractéristiques (signatures) des images-visages I_i de la base de données (Galerie) qui sont les vecteurs S_i de dimension d'' : S_1, S_2, \dots, S_M . Ce sont ces vecteurs qui doivent être mémorisés pour les exploiter dans l'étape de classification.

Il faut aussi mémoriser les matrices de projection : W , P^* et Q^* et le vecteur moyen (l'image moyenne) μ pour s'en servir dans la phase de reconnaissance.

5.2.2.2. Phase de reconnaissance (Test) :

Dans cette phase de reconnaissance (test), nous extrayons de l'image de visage détectée et normalisée I_t de la personne à reconnaître le vecteur caractéristique (signature) S_t .

La détermination du vecteur S_t est obtenue par la soumission de l'image I_t , après sa transformation en vecteur colonne X_t et ensuite en vecteur-visage centré \bar{X}_t ($\bar{X}_t = X_t - \mu$), aux trois projections successives suivantes :

Projection 1 : $Y_t = W^T \cdot \bar{X}_t \quad (5.14)$

Projection 2 : $Z_t = P^{*T} \cdot Y_t \quad (5.15)$

Projection 3 : $S_t = Q^{*T} \cdot Z_t \quad (5.16)$

5.2.3. Discussion

L'approche **ACP-ADLdouble** proposée et décrite en détail ci-dessus est une méthode de nature globale qui ne nécessite aucune connaissance a priori sur l'image de visage à reconnaître et se révèle très efficace en réduction de dimensionnalité tout en conservant beaucoup d'information utile. Cette approche lors de son test sur la base de données AT&T, nous a permis d'obtenir des performances très satisfaisantes en termes de rapidité dans la phase de reconnaissance et du taux de reconnaissance. Elle tolère les variations dues aux occultations partielles et le changement d'échelle. En revanche, lorsqu'elle est testée sur des images de visages qui possèdent des fortes variations d'illumination ou de pose, en particulier, à partir d'un angle de $\pm 25^\circ$, le taux de reconnaissance chute. Cette dégradation s'explique par le fait que l'approche est de nature globale. En effet, lorsqu'il y a une forte variation de pose ou d'illumination, il y a une perte d'information puisqu'on traite l'image de visage dans sa totalité (voir figure 5.2). En conséquence, il y a un manque de correspondance entre deux

images de visage de la même personne. Dans un tel cas, l'amplitude des variations inter-classe (entre deux images de personnes différentes) est souvent moins grande que celle intra-classe (entre deux images d'une même personne présentant des poses ou des illuminations différentes). Ce problème donne des faux résultats dans la phase de classification.

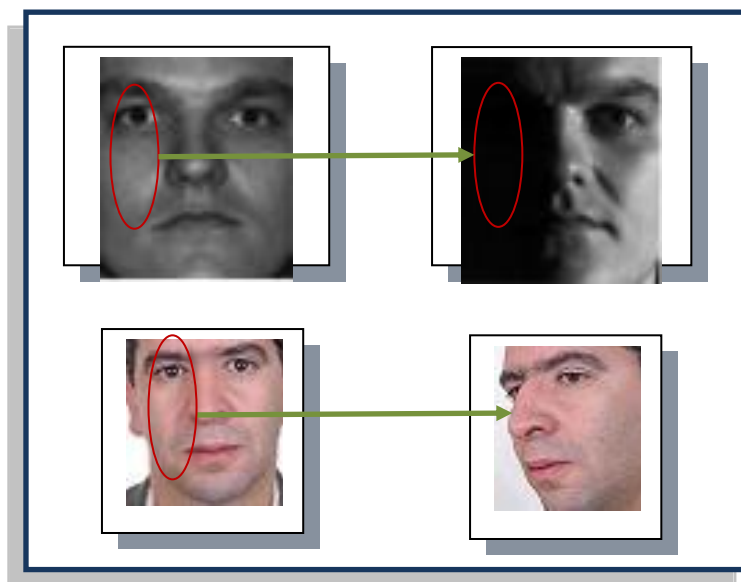


Fig. 5.2. Deux exemples montrant le manque de correspondance en raison de la perte d'information due à la forte variation d'illumination (en haut) ou la variation de pose (en bas)

5.3. Deuxième approche d'extraction des caractéristiques : Descripteur de Gabor, Estimateur de pose et ACP-ADL double

5.3.1. Principe de base de l'approche

Cette deuxième approche est élaborée dans le sens de surmonter les problèmes de la première approche, à savoir : la chute du taux de reconnaissance lorsque nous présentons au système des images de visage ayant des fortes variations de pose et d'illumination. Pour atteindre cet objectif, nous avons apporté à la première approche deux améliorations.

La première amélioration consiste dans l'introduction du descripteur de Gabor. Au lieu de soumettre l'image de visage directement à la projection *ACP-ADLdouble*, nous procédons tout d'abord à une modélisation de l'image de visage avec les ondelettes de Gabor (filtres de Gabor). Cette intégration du descripteur de Gabor est justifiée par le pouvoir des ondelettes de Gabor dans la minimisation des effets liés aux variations d'illumination, d'expression et de pose qui ne dépasse pas $\pm 40^\circ$. Cette propriété est déjà prouvée par les travaux signalés dans la section 4.6.3 du chapitre précédent et est affirmée expérimentalement par l'augmentation remarquable du taux de reconnaissance lors des tests qui seront exposés dans le chapitre 7.

La deuxième amélioration proposée est faite dans le but de tolérer les fortes variations de pose. Elle consiste dans l'utilisation de l'information de pose comme paramètre d'entrée dans la phase de classification des vecteurs caractéristiques. Pour cette raison, nous avons intégré un estimateur de pose entre la phase d'extraction et celle de classification pour faciliter la reconnaissance de visages pris dans des poses variables. Cet estimateur permet d'estimer la pose du visage à reconnaître pour cibler le classificateur associé à la pose de visage en question. Par ailleurs, avec cette nouvelle approche d'extraction, nous n'utiliserons pas uniquement un seul classificateur, mais une multitude de classificateurs paramétrés par l'orientation des visages.

Les détails de cette approche seront exposés dans les sections suivantes après une présentation des outils intégrés, à savoir : le descripteur de Gabor et l'estimateur de pose.

5.3.2. Descripteur de Gabor

Nous avons évoqué dans le chapitre précédent le grand intérêt du traitement des images de visage par des ondelettes (filtres) de Gabor. Nous les avons exploitées comme un outil très efficace dans la vérification de la présence et la localisation des traits de visage (les yeux et la bouche) dans notre approche de détection de visages. Dans cette deuxième approche d'extraction des caractéristiques, nous utilisons les ondelettes (filtres) de Gabor pour concevoir un descripteur de visage.

La modélisation d'un visage par le descripteur de Gabor se réalise à l'aide du processus illustré par la figure 5.3. L'image de visage de la personne à reconnaître I_t est soumise au départ à un banc de filtres de Gabor (40 filtres de Gabor avec 8 orientations et 5 fréquences centrales). La construction de ce banc de filtres est déjà traitée dans la section 4.6.3 du chapitre précédent. Les 40 réponses issues des filtres sont ensuite transformées en 40 vecteurs $G_{tk} \{k=1, \dots, 40\}$ et concaténés par la suite dans un seul vecteur nommé G_t ($G_t = (G_{t1}, G_{t2}, \dots, G_{t40})$). C'est ce vecteur G_t résultant qui représente le descripteur de Gabor. Vu que ce vecteur possède une taille très grande ($40 \times l \times h$, avec (l, h) est la taille de la fenêtre d'un filtre de Gabor), il doit subir une projection dans un sous-espace *ACP-ADLdouble* pour réduire sa dimension. Finalement, le vecteur S_t obtenu après projection, représente le vecteur caractéristique (signature) associé au visage à reconnaître.

De même, le sous-espace de projection *ACP-ADL double* doit être construit dans la phase d'enrôlement à partir des descripteurs de Gabor $G_i \{i= 1, \dots, M\}$ des images de visage de la base de données à l'aide du même processus.

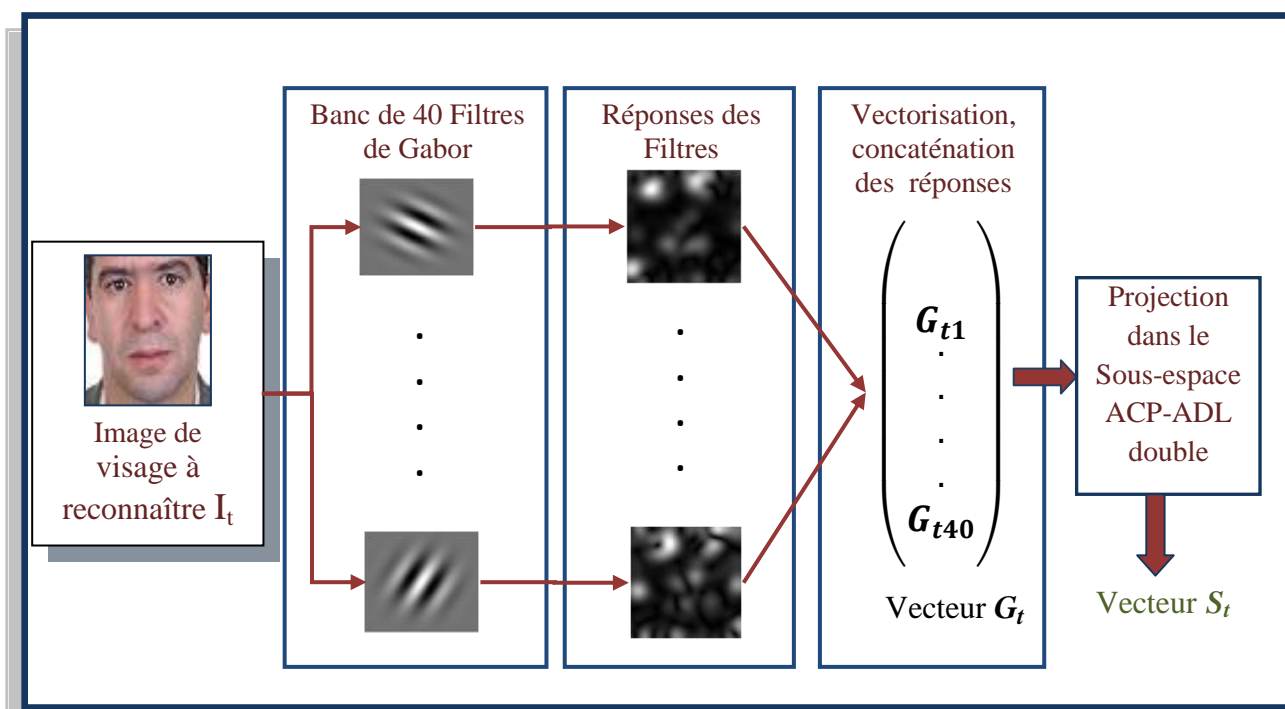


Fig. 5.3. Processus d'extraction du vecteur caractéristique par le descripteur de Gabor

5.3.3. Estimateur de pose

5.3.3.1. Présentation de l'estimation automatique de pose

Le but d'un estimateur de pose est l'estimation automatique de la pose du visage (de la tête) dans des images ou des vidéos. Géométriquement parlant, cela consiste à définir un repère associé au visage et à estimer sa rotation par rapport à un repère global de référence. Ce dernier pourra par exemple être associé au corps et défini par les plans: sagittal, frontal et transversal du corps (figure 5.4). On pourra également choisir un repère indépendant du corps: il peut s'agir de la position de la caméra de capture comme repère de référence (figure 5.4) ou d'une position de référence du visage (l'image de face, par exemple). On trouve couramment une dernière représentation qui consiste à définir uniquement une direction dans l'espace à l'aide des angles de rotation de la tête : pan, roll et tilt (figure 5.4). Dans notre cas, nous prenons le plan frontal du visage comme référence (figure 5.5).

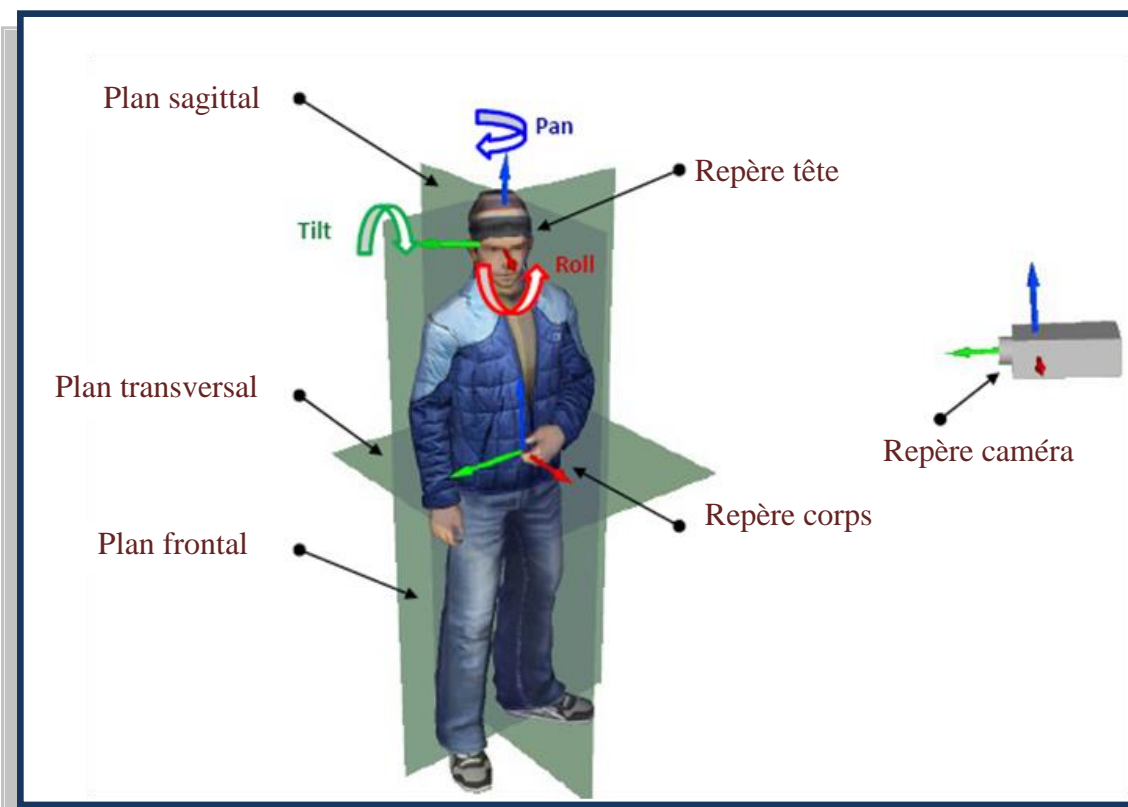


Fig. 5.4. Illustration de quelques repères de références utilisés dans l'estimation de pose : Repère corps, angles de rotation et repère caméra.

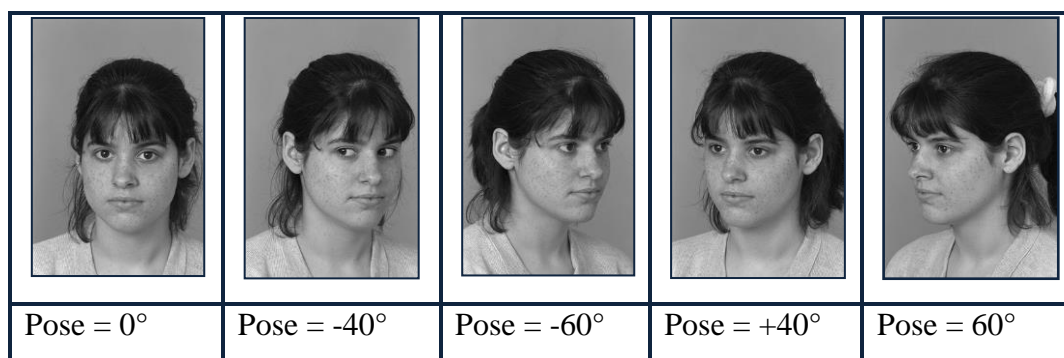


Fig. 5.5. Paramétrage de la pose du visage par rapport au plan frontal du visage comme référence

L'estimation de pose constitue un maillon essentiel dans la chaîne de traitement de nombreuses applications d'interface homme-machine. C'est pour cette raison qu'on trouve actuellement une multitude de méthodes réalisant cette fonction. Ces méthodes sont de nature globale si elles utilisent globalement l'image d'un visage pour en déduire sa pose, ou géométriques si elles infèrent la pose à partir des positions relatives de certains éléments du visage tels que les yeux et la bouche. Par ailleurs, il existe des approches qui estiment la pose par alignement d'un modèle 2D ou 3D déformable.

Pour avoir une idée approfondie sur ces méthodes, le lecteur peut se référer à l'état de l'art de [149] qui étudie un nombre important de méthodes d'estimation de la pose développées dans ces dernières années.

5.3.3.2. Estimateur de pose proposé

Avant d'entamer la présentation détaillée de notre méthode proposée pour l'estimation de pose, nous signalons que l'objectif visé est d'estimer d'une façon discrète la pose et de donner la valeur la plus proche de celle du visage parmi les valeurs suivantes: 0° , 30° , 60° , 90° , -30° , -60° , -90° et non pas la valeur exacte de la pose réelle du visage. En effet, l'estimation de pose est introduite dans notre travail dans le but de faciliter la classification, il n'est donc pas utile de connaître l'angle de la pose au degré près étant donné que le descripteur de Gabor tolère déjà les variations de pose qui ne dépassent pas $\pm 40^\circ$.

L'idée principale de conception de notre estimateur de pose se base sur la construction d'un classificateur par apprentissage supervisé en utilisant les séparateurs à vastes marges (SVM). Cette méthode de classification sera détaillée dans le chapitre suivant. Pour assurer l'apprentissage de ce classificateur, nous collectons tout d'abord quelques images de visages (les images types) de la base de données de telle sorte à former 7 groupes. Chaque groupe contient des images de visage qui ont des orientations les plus proches pour définir finalement 7 classes {classe 0° , classe 30° , ..., classe 90° }. Ensuite, ces images subissent un traitement pour en extraire un descripteur de pose pour chacune. Et ce sont donc ces descripteurs, après leur réduction qui vont être les stimulés d'apprentissage de notre classificateur SVM. Lors de la présentation d'une image de visage dont la pose à estimer (ce visage n'est pas forcément disponible au moment de l'apprentissage du classificateur SVM) à l'entrée du classificateur, ce dernier déterminera son classe d'orientation d'appartenance à partir de son descripteur de pose (figure 5.6). Autrement dit, la classe qui offre un score de ressemblance le plus élevé.

Dans la littérature, les descripteurs exploités pour estimer la pose, sont généralement inspirés des outils utilisés dans l'analyse des visages ou des formes, tels que : les filtres de Sobel ou Canny qui sont utilisés pour extraire les contours du visage, les filtres de Gabor, l'opérateur non linéaire Local Binary Patterns (LBP), les descripteurs de forme (les descripteurs de Fourier, les moments géométriques...), Les histogrammes de couleur, etc....

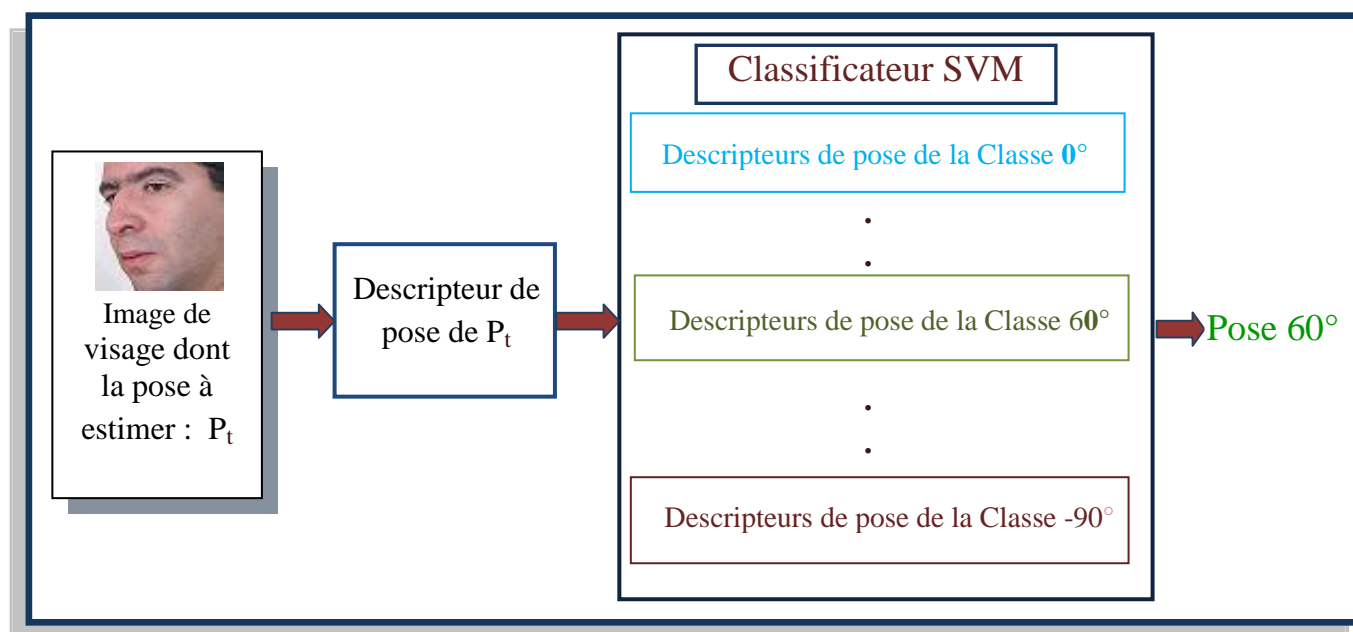


Fig. 5.6. Exemple d'illustration du principe de fonctionnement de l'estimateur de pose proposé

Concernant notre estimateur de pose, nous avons opté pour les filtres de Gabor puisque nous les utilisons en avance comme descripteur de visage. Or, dans la définition du descripteur de pose, nous n'allons pas prendre en considération toutes les réponses des 40 filtres comme le cas du vecteur descripteur de Gabor G_i d'un visage (déjà expliqué dans la section 5.3.3), mais nous ne sélectionnerons que les réponses qui portent une information utile pour l'estimation de la pose. En effet, dans l'estimation de pose nous cherchons à discriminer les poses des visages et non pas les visages elles mêmes. C'est pour cette raison que nous avons procédé à une multitude d'opérations de filtrage de plusieurs images de visage à différentes pose. Ce travail est fait dans le but de sélectionner les paramètres des filtres (la fréquence centrale et l'angle d'orientation du filtre) qui permettent de caractériser mieux la pose d'un visage.

Le résultat d'un tel travail expérimental montre que les filtres de Gabor ayant respectivement la fréquence centrale (u) et l'angle d'orientation (v) $\{(1,4) ; (2,4) ; (3,4)\}$ sont capables de caractériser l'information de pose d'un visage d'une manière efficace.

5.3.4. Description détaillée de la deuxième approche d'extraction

En combinant le descripteur de Gabor, l'*ACP-ADL double* et l'estimateur de pose développés et présentés dans les sections précédentes, nous obtenons une nouvelle approche d'extraction de nature hybride. D'une part, nous extrayons à partir de l'image globale de

visage un vecteur caractéristique contenant l'information la plus pertinente modélisée par le descripteur de Gabor et la plus réduite par *l'ACP-ADLdouble*. D'autre part, nous ciblons l'information de pose générée par l'estimateur de pose qui va résoudre au moment de la classification, le problème de chevauchement inter-classe.

Dans cette approche illustrée par le schéma synoptique de la figure 5.7, pour reconnaître le visage d'une personne qui se présente devant le système, le processus à exécuter suit les étapes suivantes :

- L'image de visage détecté et localisé I_t est soumise à l'entrée du descripteur de Gabor qui permet de retourner deux vecteurs descripteurs :
 - Le vecteur G_t concaténant toutes les réponses des 40 filtres. C'est le vecteur descripteur du visage.
 - Le vecteur G_{tp} concaténant les réponses des filtres comme il est expliqué dans la section de l'estimateur de pose. C'est le vecteur descripteur de la pose du visage.
- Le vecteur G_{tp} est ensuite soumis à l'entrée de l'estimateur de pose. Ce dernier donne comme résultat une valeur reflétant la classe de la pose la plus proche à la pose réelle du visage.
- En fonction de cette valeur de pose, le vecteur G_t est commuté vers le sous-espace *ACP-ADLdouble* associé à la classe de pose en question dans le but de réduire la taille du vecteur G_t . Après projection de cette dernière, nous obtenons le vecteur caractéristique S_t
- Et finalement, ce vecteur S_t subit une classification au sein du classificateur spécifique à la pose du visage pour prendre une décision sur son identité.

Pour terminer cette présentation de cette nouvelle approche d'extraction, il faut noter que les 7 sous-espaces de projection *ACP-ADL double* et les 7 classificateurs sont construits dans la phase d'enrôlement à partir des descripteurs de Gabor $G_i \{i= 1, \dots, M\}$ des images de visage de la base de données. Ceci est bien sûr après répartition de ces images en 7 Galeries paramétrées par l'information de pose.

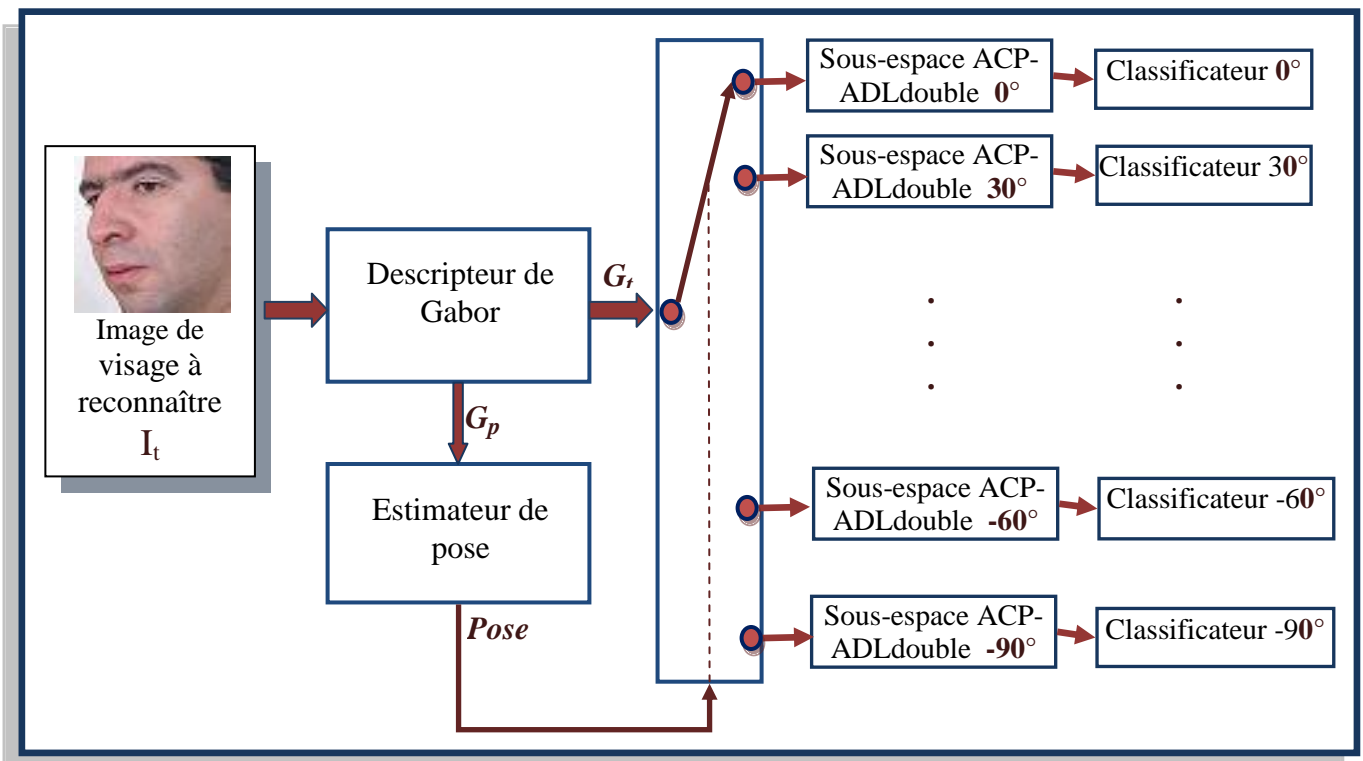


Fig. 5.7. Schéma synoptique de la deuxième approche d'extraction des caractéristiques

5.4. Conclusion

Dans ce chapitre nous avons proposé deux approches d'extraction des caractéristiques. La première est l'*ACP-ADL double*, une méthode de nature globale, qui se révèle très efficace en réduction de dimensionnalité tout en garantissant un temps très court en phase de reconnaissance et un taux de reconnaissance très raisonnable dans un environnement contrôlé. Cependant, cette dernière performance se dégrade à partir du moment où l'éclairage ou la pose du visage à reconnaître présente une variation importante. Face à ce problème, nous sommes tournés vers une autre approche dans laquelle nous avons préservé les avantages de la première en apportant d'intéressantes améliorations. En conséquence, nous avons abouti à une nouvelle approche d'extraction des caractéristiques de nature hybride. D'une part, nous extrayons à partir de l'image globale de visage un vecteur caractéristique contenant l'information la plus pertinente modélisée par le descripteur de Gabor et la plus réduite par l'*ACP-ADL double*. D'autre part, nous ciblons l'information de pose générée par l'estimateur de pose. Ce dernier est conçu de manière à résoudre au moment de la classification le problème de chevauchement inter-classe et intra-classe causé par les fortes variations de pose.

Dans le prochain chapitre, nous nous intéresserons à l'étude des méthodes de classification et de décision. Nous nous focaliserons plus sur les méthodes les plus utilisées en reconnaissance de visages.

Chapitre 6

Classification et Décision

6.1. Introduction

Jusqu'à maintenant, nous avons présenté des approches qui effectuent la détection et la localisation de visage, ainsi que l'extraction des caractéristiques (signatures). Le traitement de ces étapes est semblable pour les deux modes de fonctionnement (identification ou authentification).

Par contre, les deux étapes de classification et de décision que nous allons étudier se distinguent dans leur traitement d'un mode à l'autre. Ces deux étapes permettent de classifier le vecteur caractéristique (signature) du visage d'une personne se présentant devant le système et fournir la décision finale sur son identification ou son authentification. Les approches proposées dans la littérature pour résoudre ce problème relèvent de la classification automatique de données, champ de recherche largement exploré durant les dernières décennies pour de nombreuses applications. Des volumes entiers ont été publiés sur le sujet.

La classification (clustering en anglais) est une méthode mathématique d'analyse de données qui permet de les regrouper en plusieurs classes de telle sorte que les données d'une même classe soient le plus semblables possible et que les classes soient le plus distinctes possibles. En effet, cette méthode doit être capable de modéliser au mieux les frontières qui séparent les classes les unes des autres. Cette modélisation fait appel à la notion de fonction discriminante.

Le champ des méthodes en classification automatique est très vaste au premier abord. Le choix d'une parmi elles nécessite leur synthèse afin d'opter pour la plus adaptée à l'application envisagée. Ce choix est souvent expliqué et argumenté par :

- Le type de l'algorithme utilisé lors de l'étape d'extraction des caractéristiques.
- Le volume de la base de données (Galerie) d'enrôlement.
- La taille des vecteurs caractéristiques.
- Le temps de réponse de cette méthode de classification. En effet, ce temps doit être minimal puisque cette étape se fait en phase de reconnaissance (en ligne).

Ce chapitre est consacré à l'étude des méthodes de classification les plus utilisées en reconnaissance de visage, à savoir : les distances métriques, les réseaux de neurones et les Séparateurs à Vastes Marges (Supports Vecteurs Machines, SVM). Cette étude sera suivie par une synthèse expérimentale dans le chapitre suivant.

6.2. Distances métriques

Lorsqu'on souhaite comparer deux vecteurs caractéristiques issus de l'étape d'extraction des caractéristiques, on peut effectuer une mesure du degré de divergence (différence) entre ces deux vecteurs en utilisant une distance métrique.

Nous présentons dans cette section les définitions des deux catégories de distance métrique les plus utilisées en reconnaissance de visages. La première est constituée des distances Euclidiennes, tandis que la deuxième concerne les distances de Mahalanobis. Ensuite, nous décrivons comment exploiter ces distances dans les étapes de classification et de décision pour un système de reconnaissance de visages.

6.2.1. Définition d'une distance métrique

Une distance métrique est une métrique qui décrit la distance entre deux points d'un ensemble E . Cette distance est une application qui respecte les propriétés suivantes $\forall i, j, k \in E$:

$$1) \quad D(i, j) = D(j, i) \quad (6.1)$$

$$2) \quad D(i, j) \geq 0 \quad (6.2)$$

$$3) \quad D(i, j) = 0 \Leftrightarrow i = j \quad (6.3)$$

$$4) \quad D(i, j) \leq D(i, k) + D(k, j) \quad (6.4)$$

6.2.2. Les distances Euclidiennes

Les distances Euclidiennes se calculent directement dans l'espace original des vecteurs caractéristiques. Elles sont définies à partir de la distance de Minkowski d'ordre p dans un espace Euclidien \mathbf{R}^n (n est la dimension de l'espace Euclidien et des vecteurs caractéristiques).

Considérons deux vecteurs caractéristiques : $S = (S_1, S_2, \dots, S_n)$ et $S_t = (S_{t1}, S_{t2}, \dots, S_{tm})$, la distance de Minkowski d'ordre p entre ces deux vecteurs est définie par :

$$D_p(S, S_t) = \left(\sum_{k=1}^n |S_k - S_{tk}|^p \right)^{1/p} \quad (6.5)$$

C'est à partir de cette formule générique que vont être définies des distances couramment utilisées dans les algorithmes de reconnaissance du visage.

Pour $\mathbf{p} = 1$, on obtient la **distance de Manhattan** (aussi appelée distance « city-block » ou métrique absolue) :

$$D_1(S, S_t) = \sum_{k=1}^n |S_k - S_{tk}| \quad (6.6)$$

Pour $\mathbf{p} = 2$, on obtient la **distance Euclidienne**, qui définit l'espace cartésien:

$$D_2(S, S_t) = \sqrt{\sum_{k=1}^n (S_k - S_{tk})^2} \quad (6.7)$$

6.2.3. La distance de Mahalanobis

La distance de Mahalanobis est une mesure de distance qui permet de déterminer la similarité entre un vecteur et un ensemble de vecteurs. Elle diffère de la distance euclidienne par le fait qu'elle prend en compte la variance et la corrélation de l'ensemble des vecteurs. Ainsi, à la différence de la distance euclidienne où toutes les composantes des vecteurs sont traitées indépendamment et de la même façon, la distance de Mahalanobis accorde un poids moins important aux composantes les plus dispersées.

En pratique, pour calculer la distance de Mahalanobis d'un vecteur caractéristique S_t et un ensemble de vecteurs caractéristiques Es , on est amené à utiliser des statistiques sur l'ensemble Es , à savoir sa moyenne μ et sa matrice de covariance C . Dans ce cas, la distance de Mahalanobis est définie comme suit:

$$D_{Mah}(S_t, Es) = \sqrt{(S_t - \mu).C^{-1}.(S_t - \mu)^T} \quad (6.8)$$

La distance de Mahalanobis peut aussi être définie comme étant la mesure de dissimilarité entre deux vecteurs S_t et S de même ensemble avec une matrice de covariance C :

$$D_{Mah}(S_t, S) = \sqrt{(S_t - S).C^{-1}.(S_t - S)^T} \quad (6.9)$$

Si la matrice de covariance est diagonale, on obtient la distance euclidienne normalisée :

$$D_{EucNorm}(S_t, S) = \sqrt{\sum_{k=1}^n \frac{(S_{tk} - S_k)^2}{\sigma_k^2}} \quad (6.10)$$

Avec : σ_k^2 est la variance du $k^{ème}$ élément du vecteur S_t , calculée pour tous les $k^{èmes}$ éléments des vecteurs de l'ensemble Es .

Si la matrice de covariance est la matrice identité, cette distance est simplement la distance euclidienne.

6.2.4. Distances métriques en reconnaissance de visages

Une fois que le type de la distance métrique est choisi, son exploitation dans les deux étapes de classification et de décision dépend de l'application envisagée (authentification ou identification).

Pour l'authentification, ces deux étapes peuvent être formalisées comme suit :

Soit S_t le vecteur caractéristique (signature) extraite de la personne à authentifier lorsqu'il se présente devant le système, et S_i le vecteur caractéristique de cette même personne enregistré dans la phase d'enrôlement. Le système doit prendre une décision en déclarant cette personne comme étant l'utilisateur légitime ou un imposteur. Pour le faire, nous déterminons la valeur booléenne de la fonction f suivante:

$$f(S_t, S_i) = \begin{cases} 1 & \text{si } D(S_t, S_i) \leq \text{Seuil} \\ 0 & \text{sinon} \end{cases} \quad (6.11)$$

Où D est la distance métrique adoptée, Seuil est la valeur au-dessous de laquelle les deux vecteurs et donc les identités sont déclarés correspondants et les valeurs 1 et 0 que peut prendre la fonction f correspondent respectivement à un l'utilisateur légitime et un imposteur.

Pour l'identification, le système de reconnaissance doit trouver l'identité d'une personne se présentant devant lui, parmi celles d'une base de données contenant des personnes déjà enrôlées, et renvoyer l'identité correspondant à cette personne ou l'identité « inconnue » si cette personne ne fait pas partie de la base. Cette opération peut être formalisée de la manière suivante :

Soit S_t le vecteur caractéristique (signature) extraite de la personne à identifier. L'identification revient à déterminer son identité Id_t parmi les identités Id_k des individus préalablement enrôlés dans le système, ou Id_0 dans le cas d'une identité inconnue. La fonction d'identification f peut ainsi être définie :

$$f(S_t) = \begin{cases} Id_k & \text{si } \min_{1 \leq k \leq M} d(S_t, S_k) \leq \text{Seuil} \\ Id_0 & \text{sinon} \end{cases} \quad (6.12)$$

Où S_k est le vecteur caractéristiques correspondant à l'identité Id_k , d est la distance choisie et Seuil est la valeur au-dessous de laquelle la personne est identifiée comme client.

6.3. Les réseaux de neurones

6.3.1. Définitions

La reconnaissance du fait que le cerveau fonctionne de manière entièrement différente de celle d'un ordinateur conventionnel a joué un rôle très important dans le développement des réseaux de neurones artificiels. Les travaux effectués pour essayer de comprendre le comportement du cerveau humain ont mené à représenter celui-ci par un ensemble de composants structurels appelés neurones, massivement interconnectés entre eux. Le cerveau humain en contiendrait plusieurs centaines de milliards, et chacun de ceux-ci serait, en moyenne, connecté à dix mille autres. Le cerveau est capable d'organiser ces neurones, selon un assemblage complexe, non-linéaire et extrêmement parallèle, de manière à pouvoir accomplir des tâches très élaborées. Par exemple, n'importe qui est capable de reconnaître des visages, alors que c'est là une tâche quasiment impossible pour un ordinateur classique. C'est la tentative de donner à l'ordinateur les qualités de perception du cerveau humain qui a conduit à une modélisation électrique de celui-ci.

Le terme *réseaux de neurones* regroupe un ensemble de méthodes d'apprentissage qui font partie d'une branche de l'informatique fondamentale sous l'appellation d'Intelligence Artificielle.

Un réseau de neurone est l'association en un graphe plus ou moins complexe d'objets élémentaires appelés neurones formels (figure 6.1). Ces derniers possèdent une fonction d'activation qui permet d'influencer les autres neurones du réseau. Les connexions entre les neurones, que l'on nomme liens synaptiques, propagent l'activité des neurones avec une pondération caractéristique de la connexion.

6.3.2. Le neurone formel

En se basant sur l'analogie, le neurone formel artificiel a été introduit pour tenter de reproduire le fonctionnement d'un neurone biologique. La première étude systématique du neurone artificiel est due au neuropsychiatre McCulloch et au logicien Pitts qui s'inspirant de leurs travaux sur les neurones biologiques et proposèrent en 1943 le premier modèle d'un neurone formel artificiel.

De façon très réductrice, un neurone biologique (figure 6.2) est une cellule qui se caractérise par :

- des dendrites; les entrées principales du neurone qui captent l'information venant d'autres neurones.

- le noyau de la cellule nerveuse; c'est en cet endroit que prend naissance l'influx nerveux, qui représente l'état d'activité du neurone;
- l'axone ; une longue fibre nerveuse qui se ramifie à son extrémité; c'est la sortie du neurone et le support de l'information vers les autres neurones;
- la synapse qui communique l'information, en la pondérant par un poids synaptique, à un autre neurone; elle est essentielle dans le fonctionnement du système nerveux.

Le neurone formel (figure 6.2) est un processeur qui réalise une somme pondérée par les poids synaptiques W_i des signaux qui lui parviennent de ses entrées X_i . Cette somme sert de paramètre à une fonction, souvent non-linéaire appelée fonction d'activation, qui la transforme en un nouveau signal transmis à la sortie.

Pour un neurone formel possédant n entrées, la somme pondérée est calculée comme suit:

$$Y = \sum_{i=1}^n W_i \cdot X_i \quad (6.13)$$

Ensuite le neurone active sa sortie grâce à une fonction non linéaire f :

$$Z = f\left(\sum_{i=1}^n W_i \cdot X_i\right) \quad (6.14)$$

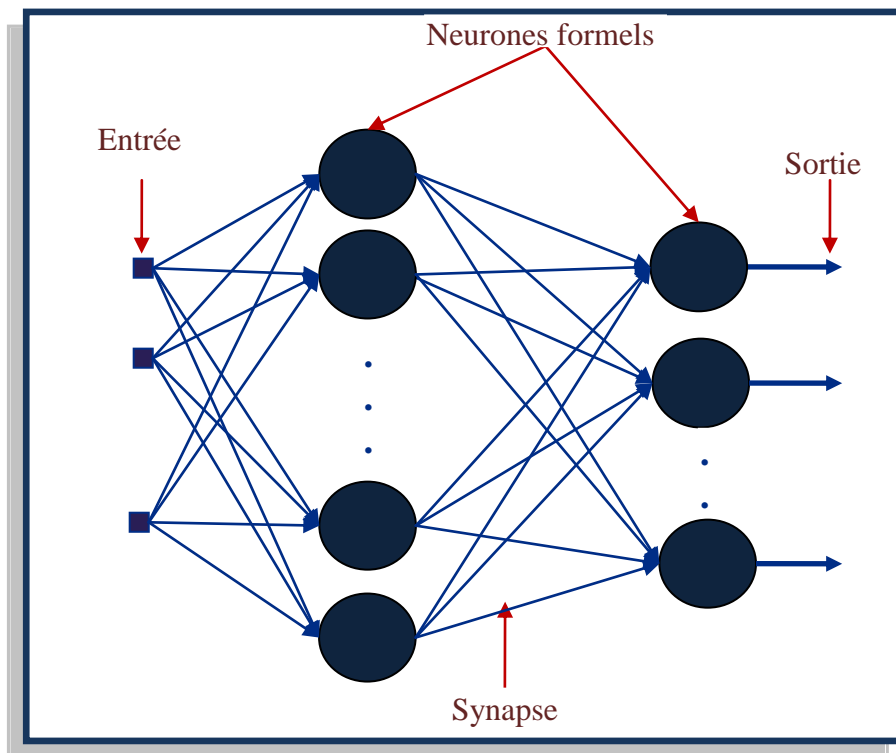


Fig. 6.1. Exemple d'architecture d'un réseau de neurone montrant les éléments de base qui le constituent

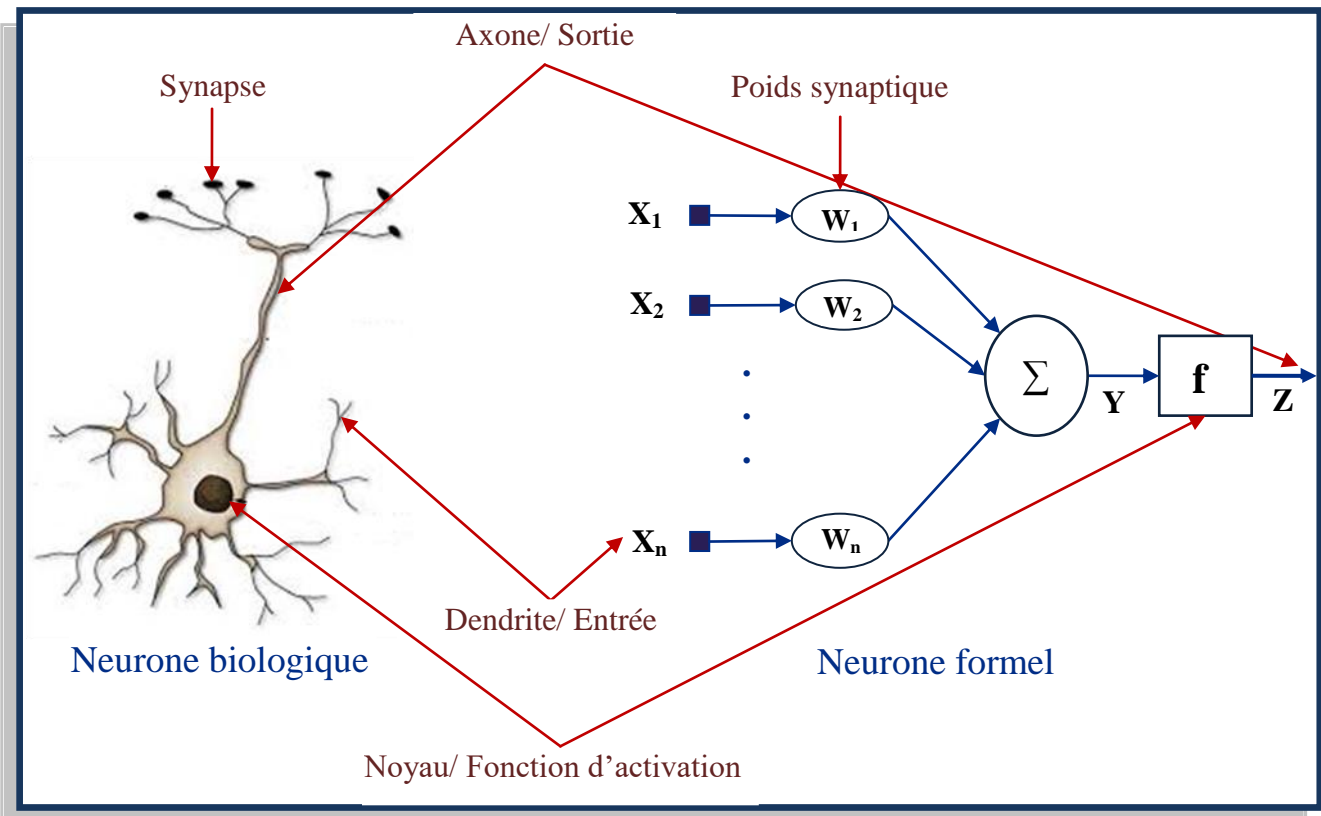


Fig. 6.2. Illustration de l'analogie entre un neurone biologique (à gauche) et un neurone formel (à droite)

Les différents types de neurones formels se distinguent par la nature de leur fonction d'activation f . Les fonctions d'activation classiques sont [150] [151]:

▪ La fonction linéaire : $f(y) = y$ (6.15)

▪ La fonction de Heavyside : $f(y) = \begin{cases} 0 & \text{si } y < 0 \\ 1 & \text{si } y \geq 0 \end{cases}$ (6.16)

▪ La fonction sigmoïde : $f(y) = \frac{1}{1 + \exp(-cy)}$ (6.17)

▪ La fonction radiale : $f(y) = \exp\left(\frac{-y^2}{2}\right)$ (6.18)

▪ La fonction tangente hyperbolique : $f(y) = \frac{1 - \exp(-cy)}{1 + \exp(-cy)}$ (6.19)

6.3.3. Applications des réseaux de neurones

Se trouvant à la croisée de différents domaines (informatique, électronique, science cognitive, neurobiologie et même philosophie), l'étude des réseaux de neurones est une voie

prometteuse pour l'Intelligence Artificielle, et possède des applications dans de nombreux domaines:

- Industrie : contrôle de qualité, diagnostic de panne, corrélations entre les données fournies par différents capteurs, analyse de signature ou d'écriture manuscrite...
- Télécommunications et informatique: analyse du signal, élimination du bruit, reconnaissance de formes, de visages et d'écriture, compression de données, classification de données...
- Finance : prévision et modélisation du marché (cours de monnaies...), sélection d'investissements, attribution de crédits...
- Environnement : évaluation des risques, analyse chimique, prévisions et modélisation météorologiques, gestion des ressources...

6.3.4. Architectures des réseaux de neurones

Devant la diversité des applications des réseaux de neurones citées en haut, les recherches ont engendré une panoplie d'architectures de réseaux dont chacune répond parfaitement à une application donnée. Ces architectures définissent la manière d'organisation des neurones. Il existe plusieurs architectures qui peuvent être classées en deux grandes catégories:

6.3.4.1. Les réseaux "FEED-FORWARD"

Appelés aussi "réseaux de type Perceptron", ce sont des réseaux dans lesquels l'information se propage de couche en couche sans retour en arrière possible.

6.3.4.1.1. Les Perceptrons

Le perceptron monocouche: C'est un réseau simple, puisque il ne se compose que d'une couche d'entrée et d'une couche de sortie. Il a été conçu dans un but premier de reconnaissance des formes. Cependant, il peut aussi être utilisé pour faire de la classification et pour résoudre des opérations logiques simples (telle "ET" ou "OU"). Sa principale limite est qu'il ne peut résoudre que des problèmes linéairement séparables. Il suit généralement un apprentissage supervisé selon la règle de correction de l'erreur (ou selon la règle de Hebb).

Le perceptron multicouche (PMC): C'est une extension du précédent, avec une ou plusieurs couches cachées entre l'entrée et la sortie. Chaque neurone dans une couche est connecté à tous les neurones de la couche précédente et de la couche suivante (excepté pour les couches d'entrée et de sortie) et il n'y a pas de connexions entre les cellules d'une même couche. La fonction d'activation utilisée dans ce type de réseaux est principalement la fonction sigmoïde. Il peut résoudre des problèmes non-linéairement séparables et des

problèmes logiques plus compliqués, et notamment le fameux problème du XOR. Il suit aussi un apprentissage supervisé selon la règle de correction de l'erreur.

6.3.4.1.2. Les réseaux à fonction radiale

Ce sont les réseaux que l'on nomme aussi RBF ("Radial Basic Functions" en anglais). L'architecture est la même que pour les PMC, cependant les fonctions de base utilisées ici sont des fonctions Gaussiennes. Les RBF seront donc employés dans les mêmes types de problèmes que les PMC particulièrement en classification et en approximation de fonctions. L'apprentissage le plus utilisé pour les RBF est le mode hybride et les règles sont soit la règle de correction de l'erreur, soit la règle d'apprentissage par compétition.

6.3.4.2. Les réseaux "FEED-BACK"

Appelés aussi "réseaux récurrents", ce sont des réseaux dans lesquels il y a un retour en arrière de l'information.

6.3.4.2.1. Les cartes auto-organisatrices de Kohonen

Ce sont des réseaux à apprentissage non-supervisé qui établissent une carte discrète, ordonnée topologiquement en fonction de patterns d'entrée. Le réseau forme ainsi une sorte de treillis dont chaque nœud est un neurone associé à un vecteur de poids. La correspondance entre chaque vecteur de poids est calculée pour chaque entrée. Par la suite, le vecteur de poids ayant la meilleure corrélation, ainsi que certains de ses voisins, vont être modifiés afin d'augmenter encore cette corrélation.

6.3.4.2.2. Les réseaux de Hopfield

Les réseaux de Hopfield sont des réseaux récurrents et entièrement connectés. Dans ce type de réseau, chaque neurone est connecté à chaque autre neurone et il n'y a aucune différenciation entre les neurones d'entrée et de sortie. Ils fonctionnent comme une mémoire associative non-linéaire et sont capables de trouver un objet stocké en fonction de représentations partielles ou bruitées. L'application principale des réseaux de Hopfield est l'entrepôt de connaissances mais aussi la résolution de problèmes d'optimisation. Le mode d'apprentissage utilisé ici est le mode non-supervisé.

6.3.4.2.3. Les réseaux ART

Les réseaux ART ("Adaptative Resonance Theorie") sont des réseaux à apprentissage par compétition. Les vecteurs de poids ne seront adaptés que si l'entrée fournie est suffisamment proche d'un prototype déjà connu par le réseau. On parlera alors de résonance. A l'inverse, si

l'entrée s'éloigne trop des prototypes existants, une nouvelle catégorie va alors se créer, avec pour prototype, l'entrée qui a engendrée sa création.

Dans ce manuscrit, nous allons nous concentrer sur le type de perceptron multicouche qui est le plus répandu pour mettre en œuvre des applications de classification.

6.3.5. Les perceptrons multicouches

La mise en cascade de perceptrons conduit à ce qu'on appelle le perceptron multicouche (*Multi-Layer Perceptron, MLP*) (figure 6.3). Lorsque le vecteur d'entrée est présenté à l'entrée du réseau, il est communiqué à tous les neurones de la première couche. Les sorties des neurones de cette couche sont alors communiquées aux neurones de la couche suivante, et ainsi de suite. La dernière couche du réseau est appelée couche de sortie, les autres étant désignées sous le terme de couches cachées car les valeurs de sortie de leurs neurones ne sont pas accessibles de l'extérieur.

La fonction d'activation des neurones doit absolument être non linéaire, car sinon, le perceptron multicouche ne ferait qu'implanter une série de transformations linéaires consécutives, qui pourraient dès lors se réduire à une seule. Grâce à l'utilisation de fonctions d'activations non linéaires, le perceptron multicouche est capable de générer des fonctions discriminantes non linéaires.

L'apprentissage du perceptron multicouches est supervisé, et consiste à adapter les poids synaptiques des neurones de manière à ce que le réseau soit capable de réaliser une transformation donnée, représentée par un ensemble d'exemples constitué d'une suite de vecteurs d'entrées associée à une autre suite de vecteurs de sorties désirées.

L'algorithme d'apprentissage du perceptron multicouche, connu sous le nom d'algorithme de **rétro-propagation du gradient**, nécessite toutefois que les fonctions d'activation des neurones soient continues et dérivables. Cette condition mise à part, celles-ci peuvent cependant être quelconques, offrant ainsi une grande liberté de modélisation.

Les détails de l'algorithme de **rétro-propagation du gradient** sont expliqués dans l'annexe B.

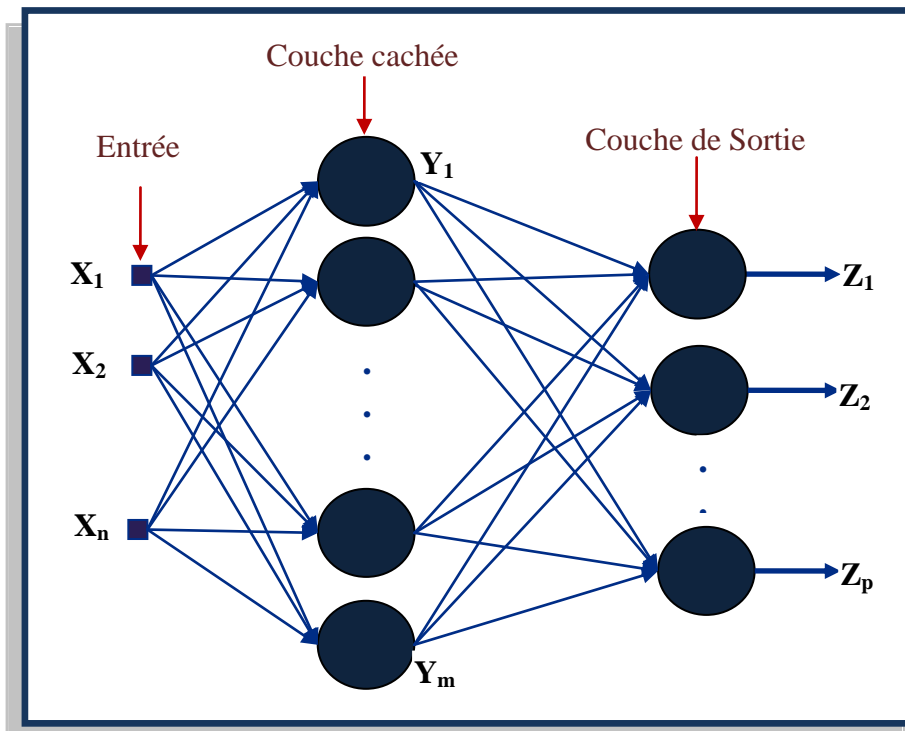


Fig. 6.3. Exemple de perceptron multicouche à une seule couche cachée

6.3.6. Réseaux de neurones en reconnaissance de visages

Les réseaux de neurones ont été employés dans tous les modules intervenant dans la chaîne de traitement d'un système automatique de reconnaissance de visages. Ils sont utilisés pour la détection et la localisation de visage (nous avons déjà cité des exemples de travaux élaborés dans ce sens dans la section 4.2.3 du chapitre 4), pour l'extraction des caractéristiques et pour la classification [152] [153] [154].

Grâce à leur capacité d'apprentissage automatique à partir de données modélisant le problème à résoudre, nous avons exploité les perceptrons multicouches comme moyen de classification des vecteurs caractéristiques extraites des visages. Pour atteindre cet objectif, les deux procédures suivantes doivent être exécutées :

- Durant la phase d'enrôlement, un apprentissage supervisé du perceptron doit être réalisé à partir de l'ensemble des vecteurs caractéristiques extraits des visages de la base de données. Pour le faire, on soumet d'une manière itérative à l'entrée du perceptron plusieurs fois et dans un ordre aléatoire ces vecteurs. Et pour chaque vecteur sélectionnée, on considère que les sorties souhaitées du perceptron sont toutes nulles sauf celle qui correspond à la classe (identité de la personne) dont fait partie ce vecteur qui doit prendre la valeur 1. On ajuste par la suite les poids des connexions du perceptron en utilisant l'algorithme de rétro-propagation.

▪ Durant la phase de reconnaissance (test), le vecteur caractéristique du visage de la personne à reconnaître est présenté à l'entrée du perceptron. La sortie de ce dernier qui prend la valeur maximale (la plus proche à 1) indique à quelle classe (identité de la personne) appartient ce visage.

Notons que les perceptrons multicouches sont généralement utilisés comme moyen de classification dans les systèmes de reconnaissance de visages qui opèrent dans le mode d'identification.

6.4. Les séparateurs à vastes marges (SVM)

6.4.1. Présentation

Les Séparateurs à Vastes Marges ou les machines à vecteurs de support (*Support Vectors Machines en anglais, SVM*) sont un ensemble de techniques d'apprentissage destinées à résoudre des problèmes de classification, c'est-à-dire décider à quelle classe appartient un échantillon, ou de régression, c'est-à-dire prédire la valeur numérique d'une variable. Les SVM ont été introduits par Vladimir Vapnik en 1995 [155], puis ont fait l'objet d'un brevet d'invention déposé en 1997.

Les SVM suscitent actuellement un vif intérêt dans la communauté de la machine learning, ce qui prouve leurs gains en popularité et en utilisation dans de nombreuses applications telles que la reconnaissance de formes (les écritures manuscrites, les visages, ...), la catégorisation de textes (classification d'e-mails, classification de pages web, ...), diagnostic médical (évaluation des risques de cancer, détection d'arythmie cardiaque, ...),

Le succès des SVM est justifié par les solides bases théoriques qui les soutiennent. Il existe en effet un lien direct entre la théorie de l'apprentissage statistique et l'algorithme d'apprentissage des SVM. De plus, les SVM sont des méthodes de classification particulièrement bien adaptées au traitement des données de très hautes dimensions.

6.4.2. Fondements de base des SVM

Pour expliquer les notions sur lesquelles se base le fondement des SVM, plaçons-nous dans le cas d'une classification binaire (les données à classer sont réparties en 2 classes).

6.4.2.1. Hyperplan

L'objectif principal d'un SVM est de trouver une frontière de décision qui sépare les points de données de deux classes différentes. Cette frontière qui est appelée un séparateur doit être un hyperplan linéaire. En général, il peut y avoir plusieurs hyperplans séparateurs possibles

entre les deux classes (en supposant le problème linéairement séparable) (figure 6.4). Cependant, nous faisons un choix particulier parmi tous les séparateurs possibles : nous cherchons un **hyperplan optimal** (*Optimal Separating Hyperplane*). Ce dernier est calculé de telle sorte que sa distance aux points les plus proches dans chaque classe soit maximale.

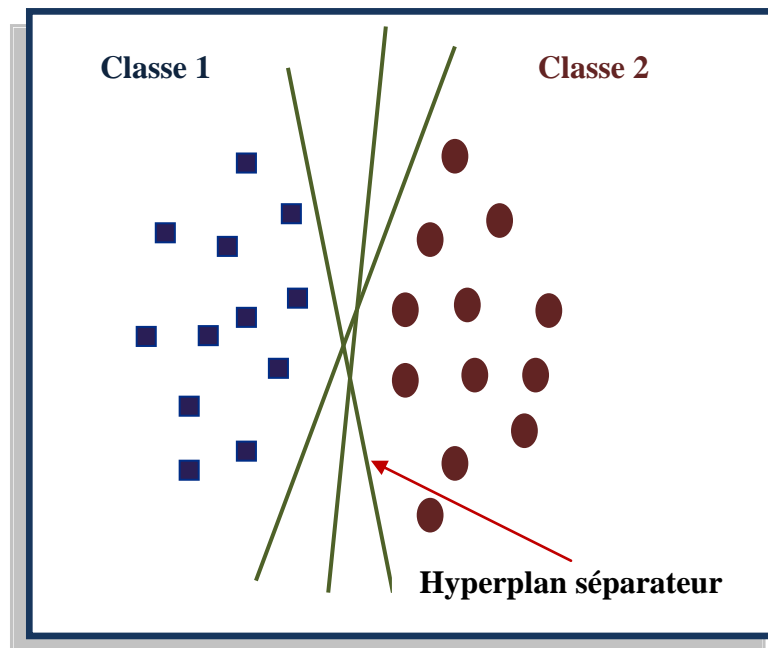


Fig. 6.4. Hyperplans séparateurs arbitraires

6.4.2.2. Vecteurs de support et marge maximale

Pour déterminer un hyperplan optimal et discriminant au sens d'un critère à fixer (maximiser la distance entre les deux classes), on utilise seulement les points de données les plus proches (les points de la frontière entre les deux classes de données) parmi l'ensemble total d'apprentissage, ces points sont appelés **vecteurs de support** (*support vectors*) (figure 6.5). La distance entre ces points est appelée **marge**. C'est cette distance qu'on souhaite maximiser (**marge maximale**).

Intuitivement, le fait d'avoir une marge plus large procure plus de sécurité lorsqu'on classe une nouvelle donnée. De plus, si l'on trouve le classificateur qui se comporte le mieux vis-à-vis des données d'apprentissage, il est clair qu'il sera aussi celui qui permettra au mieux de classer les nouveaux exemples.

6.4.2.3. Linéarité et non-linéarité

Parmi les modèles des SVM, on constate les cas linéairement séparables et les cas non linéairement séparables (figure 6.6). Les premiers sont les plus simples car ils permettent de trouver facilement le classificateur linéaire. Dans la plupart des problèmes réels il n'y a pas de

séparation linéaire possible entre les données, le classificateur de marge maximale ne peut pas être utilisé car il fonctionne seulement si les classes de données d'apprentissage sont linéairement séparables.

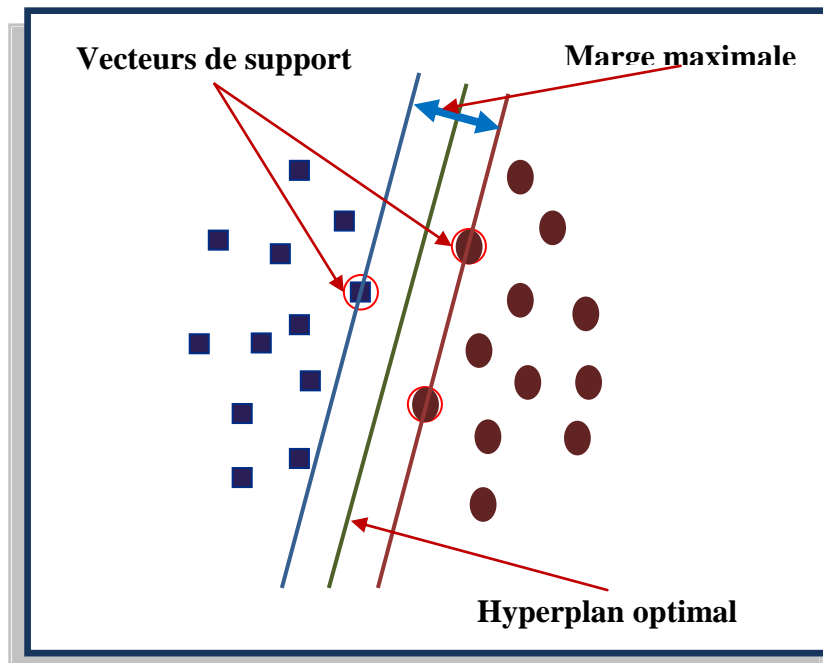


Fig. 6.5. Illustration des notions: hyperplan optimal, vecteurs de support et marge maximale

Pour surmonter le problème des cas non linéairement séparables, l'idée des SVM est de changer l'espace des données. Ce dernier peut toujours être plongé dans un espace de plus grande dimension appelé « espace de redescription ». En effet, plus la dimension de l'espace de redescription est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les données est élevée. Ceci est illustré par la figure 6.6.

La transformation de l'espace de représentation des données non-séparables linéairement à un espace de redescription est réalisée via une fonction noyau (*Kernel*). En pratique, quelques familles de fonctions noyau paramétrables sont connues et il revient à l'utilisateur de la SVM d'effectuer des tests pour déterminer celle qui convient le mieux pour son application envisagée. On peut citer les exemples de noyaux suivants : polynomiale, gaussien, sigmoïde et Laplacien.

6.4.3. SVM à plusieurs classes

Le principe du SVM expliqué dans la partie précédente se résume dans la résolution des problèmes de classification binaire, or les problèmes rencontrés dans la réalité, sont de type multi classes. D'où l'importance d'étendre le principe du SVM aux problèmes de plus de

deux classes. Il existe plusieurs types d'extensions mais nous nous contenterons d'en présenter seulement deux qui sont assez souvent utilisées. Dans la suite de ce paragraphe nous considérerons un problème à K classes avec $K > 2$.

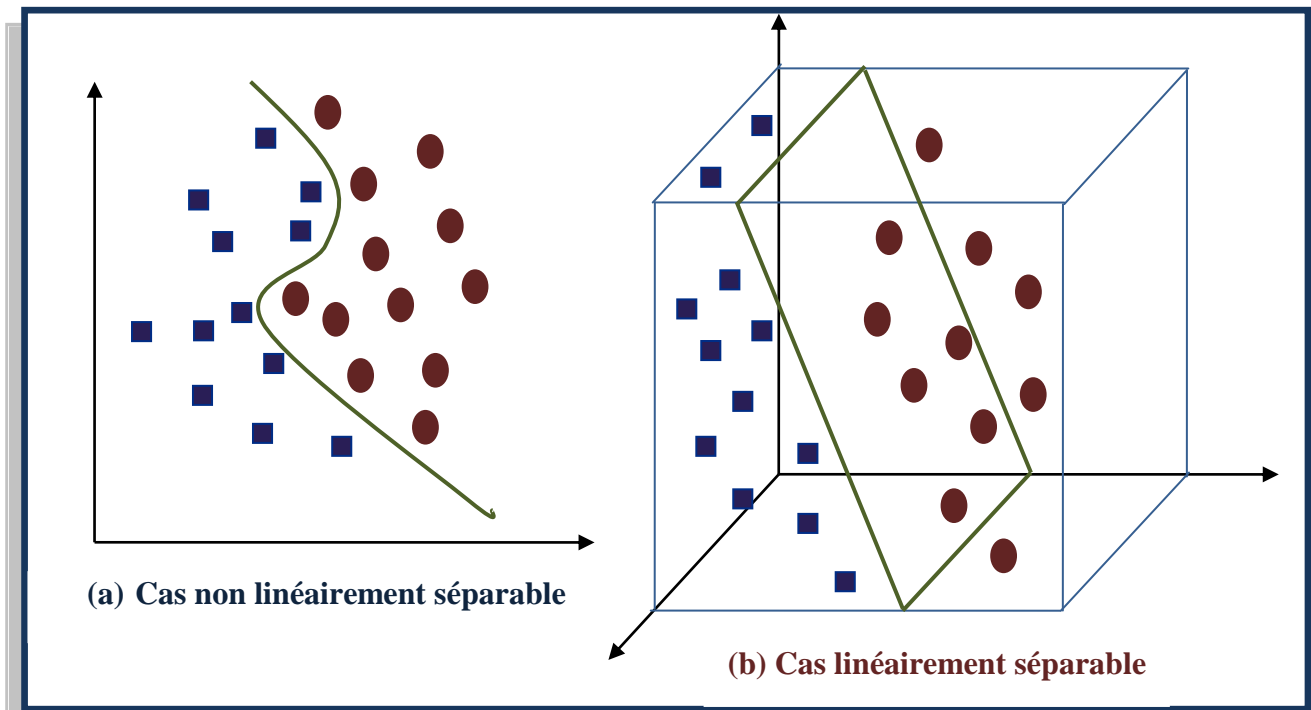


Fig. 6.6. Exemple de transformation de l'espace : (a) Cas non linéairement séparable dans un espace à deux dimensions et (b) Cas linéairement séparable dans un espace à trois dimensions

6.4.3.1. Un contre tous (One versus All)

La méthode la plus intuitive pour la gestion de la multi classes consiste à construire autant de classificateur SVM que de classe (K classificateurs). Chaque classificateur renvoie 1 si la donnée à classer appartient à la classe, -1 sinon. Il faut donc pour classer une donnée, la soumettre à tous les classificateurs, le meilleur remportant la décision. Il est évident qu'avec un nombre de classe élevé, la combinatoire peut devenir énorme. Cette méthode est appelée en anglais One-Versus-All (1-vs-A).

6.4.3.2. Un contre un (One versus One)

Il est également possible de concevoir des classificateurs spécialisés dans la comparaison classe à classe (méthode One-versus-One (1-vs-1) en anglais). Pour un problème à K classes, On a $K.(K-1)/2$ classificateurs. On soumet la donnée à classer à tous ces classificateurs 1-vs-1, la classe remportant le plus de suffrage remporte la décision.

Pour plus de détails concernant l'obtention de certaines formules mathématiques nécessaires à l'implantation des SVM pour les deux cas (linéaire ou non-linéaire) ou pour une

bonne compréhension des SVM, le lecteur pourra se référer au mémoire de la thèse de Mathieu Feuilloy [150], les pages: 31-37 et 53-56.

6.4.4. SVM en reconnaissance de visages

SVM est une méthode de classification qui a montré son efficacité dans la résolution de quelques problématiques liées au domaine de la reconnaissance de visages, telles que :

- La détection de visages dans une image, plus précisément pour la classification visage et non-visage des objets localisés dans une image (nous avons déjà cité des exemples de travaux élaborés dans ce sens au paragraphe 4.2.3 du chapitre 4).
- La détection et la localisation des traits caractéristiques du visage par la mise en cascade de plusieurs classificateurs SVM [156].
- La classification des vecteurs caractéristiques (signatures) des visages [157] [158].

Les différentes expériences ainsi que les résultats très satisfaisants obtenus par cette approche nous ont encouragés à explorer cette piste pour les deux finalités suivantes:

- La construction du classificateur utilisé pour l'estimateur de pose expliqué en détail dans le chapitre précédent. Comme nous l'avons signalé dans le chapitre précédent (paragraphe 5.3.3.2), dans cette application particulière des SVM, nous désirons discriminer 7 classes dont chacune est étiquetée par une valeur de pose parmi les valeurs : 0° , 30° , 60° , 90° , -30° , -60° , -90° . Les points de données stimulant une classe seront les vecteurs descripteurs de pose de certains exemples de visage de la base de données qui ont des poses proches dans un intervalle de 30° .

Durant la phase d'enrôlement, nous déterminons les hyperplans séparateurs des 7 classes, tandis que pour la phase de reconnaissance (Test), un visage cible est assigné à une pose selon la position de son vecteur descripteur de pose dans l'espace partitionné en régions distinctes par les hyperplans SVM.

- La classification des vecteurs caractéristiques issus de l'algorithme d'extraction des caractéristiques. De même pour cette deuxième application des SVM, nous suivons la même procédure pendant les deux phases (enrôlement et reconnaissance), sauf que dans cette situation, le nombre de classes correspond au nombre de personnes (clients) de la Galerie dont chacune est étiquetée par une identité. Les points de données de chaque classe (personne) sont les vecteurs caractéristiques des images du visage d'une même personne, prises avec plusieurs vues différentes.

6.5. Conclusion

Ce chapitre a été consacré à l'étude de trois types de méthodes de classification, les plus utilisées en reconnaissance de visage : les distances métriques, les réseaux de neurones et les séparateurs à vastes marges (SVM). Dans cette étude, nous avons présenté en premier lieu les fondements théoriques de base de chaque méthode et leurs champs d'application et ensuite, nous avons décrit les différentes étapes de leurs implémentations dans un système de reconnaissance de visage. Toutefois, pour que cette étude soit suffisamment complète, nous la finalisons par une comparaison de ces méthodes afin de cerner les avantages et les inconvénients de chaque méthode dans le contexte de la reconnaissance de visage. Cette comparaison est faite sous forme d'une synthèse expérimentale exposée dans le chapitre suivant. Ce dernier est consacré aussi à la présentation des résultats obtenus par les différentes approches proposées dans ce mémoire.

Chapitre 7

Résultats, évaluations et discussions

7.1. Introduction

Dans le cadre de l'évaluation des performances de nos approches proposées dans cette thèse, nous avons réalisé une série de tests expérimentaux. Or, pour valider cette évaluation, d'une part, nos tests sont expérimentés sur des bases de données standards, et d'autre part nous avons effectué des comparaisons avec d'autres méthodes de l'état de l'art.

Nous présentons et discutons dans ce chapitre les différents résultats obtenus. Une première section est consacrée à l'approche de détection et de localisation de visages. Tandis que la deuxième, est réservée aux deux approches d'extraction des caractéristiques et la synthèse expérimentale des méthodes de classification.

Avant d'entamer les détails de ce chapitre, notons qu'au moment de la classification et la décision, nous nous sommes placés dans le mode d'identification et exactement dans un contexte fermé, c'est à dire sous l'hypothèse forte que l'identité recherchée se trouve dans la base de données (Galerie). Nous nous sommes ainsi affranchi de la notion inhérente du seuil d'acceptabilité de l'identification présent dans un contexte ouvert. De plus, dans ce mode, nous avons la possibilité de tester les perceptrons multicouches (PMC) et les séparateurs à vastes marges (SVM) qui sont souvent exploités comme moyen de classification dans les systèmes d'identification. Cependant, toutes les approches proposées sont valides pour le cas du mode d'authentification (vérification).

7.2. La détection et la localisation de visages

Dans cette section, nous allons commencer par la description et l'analyse des résultats obtenus par le détecteur des pixels de peau développé. Ensuite, nous présentons les résultats des tests de l'approche dans sa totalité et la comparaison de ses performances avec l'algorithme de Viola et Jones. Mais avant cette présentation, nous donnons tout d'abord les caractéristiques de la base de données des images de visage sur lesquelles nous avons effectué nos expérimentations, à savoir la base : Caltech_10K_WebFaces¹⁰.

7.2.1. La base Caltech_10K_WebFaces

La base de données Caltech_10K_WebFaces contient des images de personnes recueillies à partir du Web. Les images de cette base ont été collectées et organisées par Michael Fink au

10. http://www.vision.caltech.edu/Image_Datasets/Caltech_10K_WebFaces/

sein du groupe Caltech Vision Group¹¹. La base contient en total 7092 images avec 10 524 visages. Ces visages humains ont diverses tailles et dans différents contextes, par exemple, les images de portrait, des groupes de personnes, etc. Les visages de profil ou des visages de résolution très faibles ne sont pas étiquetés. Les coordonnées des yeux, le nez et le centre de la bouche pour chaque face frontale sont fournies dans un fichier. Cette information peut être utilisée pour aligner et recadrer les visages humains ou comme une vérité de terrain pour évaluer un algorithme de détection et de localisation de visages. La résolution moyenne des images est de 304x312 pixels. Les images sont en couleur ou en niveau de gris.

7.2.2. Evaluation du détecteur de pixels de peau

Au moment du développement du modèle de la distribution de la couleur de peau que nous avons adopté par la suite pour le détecteur des pixels de peau, nous sommes amenés à effectuer des tests comparatifs sur quatre modèles de peau proposés dans la littérature. Ces modèles qui sont déjà présentés dans le paragraphe 4.4.3 du chapitre 4, utilisent des espaces de couleur différents. Les tests sont opérés sur des images couleurs de la base Caltech_10K_WebFaces et les paramètres sur lesquels nous nous sommes basés pour faire la comparaison sont :

- **Le taux de vrais positifs VP:** la probabilité qu'un pixel appartenant à la classe peau soit affecté à la classe peau.
- **Le taux de faux positifs FP :** la probabilité qu'un pixel appartenant à la classe non peau soit affecté à la classe peau.
- La capacité de faire apparaître les traits caractéristiques du visage (les yeux et la bouche) après la transformation de l'image couleur d'entrée en une image binaire (pixels peau/non-peau).

Un exemple de cette comparaison des quatre modèles avec notre modèle proposé est illustré dans la figure 7.1.

11. <http://www.vision.caltech.edu/>

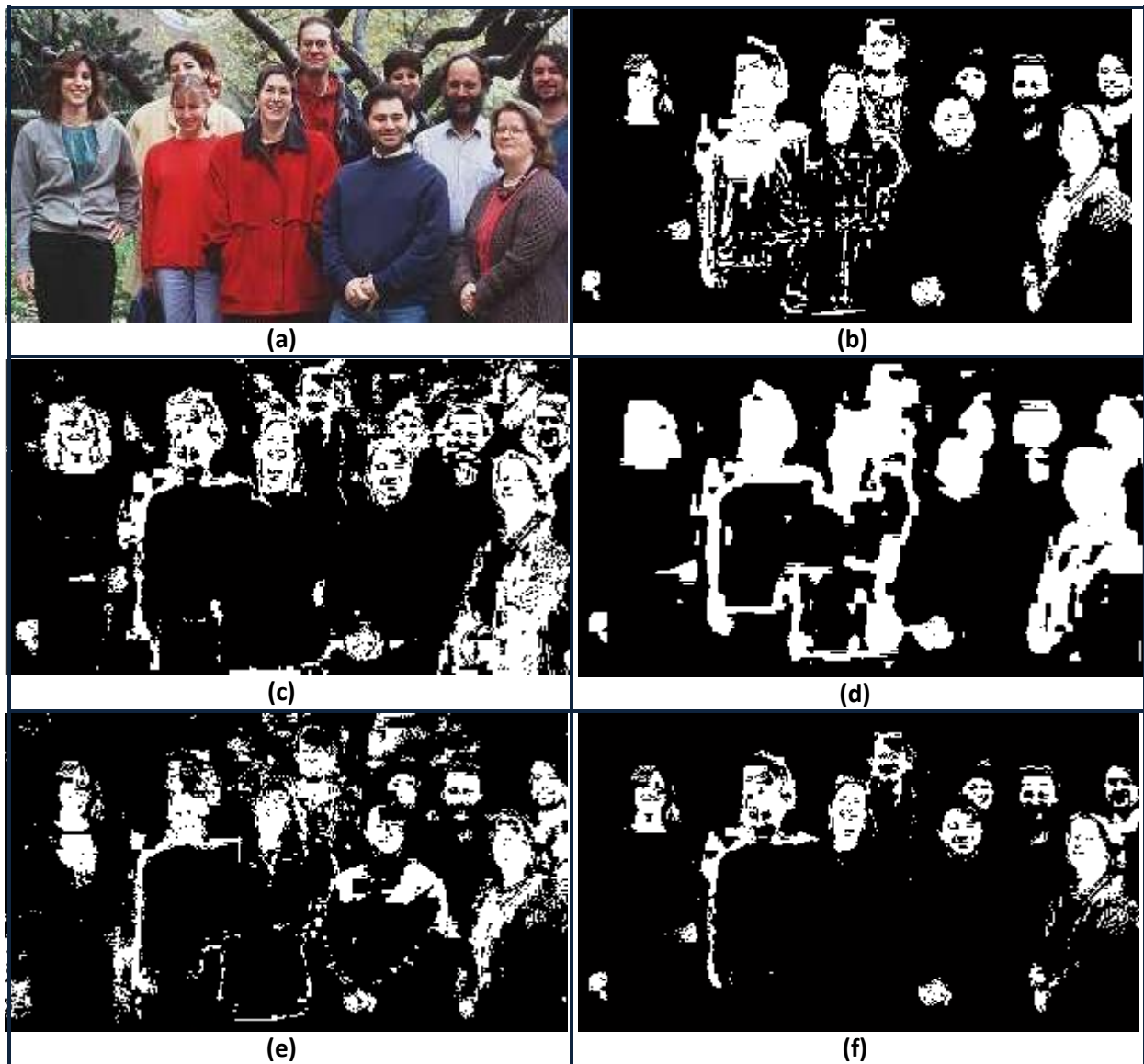


Fig. 7.1. Détection des pixels de la peau avec différents modèles : (a) image couleur originale; (b) détection à base de RGB; (c) détection à base de RGB normalisé ; (d) détection à base de YCbCr ; (e) détection à base de HSV ; (f) détection avec notre modèle de peau proposé

Discussion

La synthèse des résultats obtenus montre que le premier modèle utilisant l'espace RGB garantit un excellent taux de VP, mais sa faiblesse réside dans le taux de FP. Cette dernière est due à la fausse détection des pixels ayant des teintes qui sont proches du rouge, orange ou jaune ressemblant aux pixels de la peau. Cependant ce problème n'existe pas dans le modèle de l'espace RGB normalisé. Par contre, ce dernier détecte une partie des cheveux en tant que peau (nous remarquons dans l'image figure 7.1.c qu'il y a détection des zones des cheveux comme étant des pixels de peau). Ces deux modèles (RGB et RGB normalisé) ont un avantage commun, celui de faire apparaître les traits caractéristiques du visage. Tandis que

pour le modèle basé sur YCbCr, on détecte bien tous les pixels de peau présents dans une image (le plus grand VP), mais il y'a un grand taux de FP (même problème que le RGB normalisé avec un degré de plus). Or, ce modèle ne permet pas de laisser une trace sur les traits de visage. Alors que le plus bas score du taux VP est obtenu par le modèle basé sur HSV.

Par ailleurs, le modèle de peau proposé nous a permis d'obtenir des résultats très satisfaisants au niveau des taux VP et FP en couvrant toutes les différentes teintes que peut prendre la couleur de la peau. Ainsi, il répond à l'autre critère fixé en avance, à savoir : le modèle de peau doit être capable de faire apparaître les traits caractéristiques du visage (les yeux et la bouche). Les deux exemples illustrés dans la figure 7.2 prouvent ces deux interprétations.



Fig. 7.2. Deux exemples de détection des pixels de la peau avec notre modèle de peau proposé

7.2.3. Evaluation de l'approche de détection et de localisation de visages

Pour évaluer l'approche de détection et de localisation de visages dans sa totalité, nous avons utilisé des images couleurs de test issues de la base Caltech_10K_WebFaces et d'autres issues du Web (ces images sont ajoutées dans le but de tester la détection dans le cas des visages de profil). Ensuite, nous avons procédé à une comparaison de notre approche avec l'algorithme de Viola-Jones. Or, pour valider cette comparaison, les tests des deux méthodes doivent être menés sur le même ensemble d'images. Pour atteindre cet objectif, nous avons téléchargé une implémentation de la version 2 de l'algorithme de Viola-Jones à partir d'OpenCV¹².

Les paramètres pris en considération dans cette évaluation sont :

- Le Taux de Vraies Acceptations (*Right Acceptance Rate, RAR*) qui exprime le pourcentage des vrais visages acceptés par le système.
- Le Taux de Fausses Acceptations (*False Acceptance Rate, FAR*) qui exprime le pourcentage des régions non-visage acceptés par le système alors qu'ils devraient être rejetés.
- Le temps moyen de détection d'un visage.

Les résultats de cette comparaison, sont présentés comme suit :

- Les deux taux RAR et FAR obtenus sont exposés dans le tableau 7.1 ci-dessous. Ces taux sont relevés à partir des tests sur un ensemble de 30 images ayant différents : nombre de visages par image, taille de visage et orientation de visage.
- Le temps de détection relevé pour quelques exemples d'images (tableau 7.2).

Finalement, nous exposons à titre d'exemple, la détection et la localisation de visages dans une image couleur par les deux méthodes (figure 7.3).

Méthode	Taux RAR	Taux FAR
Notre approche	93,18 %	2,13 %
Algorithme de Viola-Jones	95,22 %	11,96 %

Tableau 7.1. Les taux RAR et FAR obtenus par notre approche de détection et l'algorithme de Viola-Jones

12. www.OpenCV.org

Nombre de visages par image	Notre approche proposée	Algorithme Viola-Jones
1	2.297s	5.048s
2	2,342s	4.095s
3	2.462s	4.703s

Tableau 7.2. Les temps de détection des deux méthodes : notre approche de détection et l'algorithme de Viola-Jones

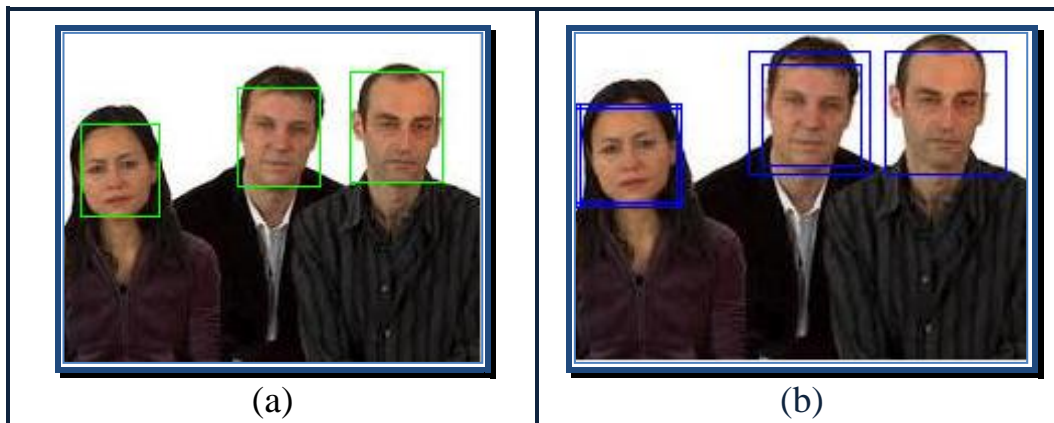


Fig. 7.3. Exemple de détection de visages (a) notre approche et (b) algorithme de Viola-Jones

Discussion

Après analyse des résultats obtenus, nous pouvons affirmer que les niveaux de fiabilité et de rapidité atteints par notre approche sont raisonnables et très satisfaisants. Et en les comparant avec ceux obtenus par l'algorithme de Viola-Jones, nous constatons que :

- L'algorithme de Viola-Jones est supérieur au niveau du taux RAR (Tableau 7.1). La cause est que notre approche rencontre des problèmes dans certains cas particuliers, comme par exemple, le cas où le visage a un éclairage non-uniforme. Dans cette situation, le détecteur des pixels de peau n'arrive pas à faire apparaître tous les traits caractéristiques (la bouche et les yeux). En conséquence, le module de classification visage et non-visage rejette ce visage.
- Toujours dans l'interprétation du taux RAR, les deux méthodes ne peuvent pas détecter des visages de profil dont la pose dépasse 60° . L'explication de la cause de ce handicap pour le cas de notre approche est que les trois traits caractéristiques n'apparaissent pas dans l'image, alors que notre modèle de visage ne traite pas ce cas.
- Le taux FAR (Tableau 7.1) constitue un point fort de notre approche. En effet, le modèle de visage adopté dans notre approche se voit sévère dans la discrimination des régions de peau au moment de la classification. De plus, le fait de réduire le champ de recherche uniquement dans les régions de peau diminue la probabilité de l'erreur. Par contre, pour

l'algorithme de Viola-Jones, le fait d'utiliser une méthode de boosting pour la classification provoque généralement une zone de confusion entre la classe de visage et la classe de non-visage. C'est pour cette raison, que la majorité des améliorations de l'algorithme de Viola-Jones proposées dans la littérature (chapitre 4) se sont concentrées sur ce problème.

- En termes de rapidité (Tableau 7.2), notre approche est la meilleure. En effet, le principe de segmentation appliqué dans notre approche réduit l'espace de recherche des visages, ce qui entraîne une augmentation de la vitesse de détection. Par contre, dans le cas de l'algorithme de Viola et Jones, on doit balayer une image à l'aide d'une fenêtre de détection de taille initiale 24px par 24px et de déterminer si un visage y est présent. Le décalage de la fenêtre se fait par un pas d'un pixel. Lorsque l'image a été parcourue entièrement, la taille de la fenêtre est augmentée par un facteur multiplicatif de 1,25. Tout ce traitement diminue la rapidité de l'algorithme.
- Le temps de détection augmente avec l'augmentation du nombre de visages dans l'image dans notre approche. Ceci est dû au traitement de classification visage et non-visage qu'on doit exécuter pour chaque région de peau segmentée. Et pour le cas de l'algorithme de Viola et Jones, ce temps augmente avec l'augmentation de la taille des visages car plus cette dernière est grande, plus la taille de la fenêtre qui va la détecter doit être grande. Or, la taille de la fenêtre est multipliée par 1,25 à chaque balayage de la totalité de l'image.

7.3. Extraction des caractéristiques et classification

Afin de renforcer l'étude des performances des approches d'extraction des vecteurs caractéristiques développées et la synthèse des méthodes de classification, nous avons exploité dans nos tests les deux bases de données de visages : AT&T et Color FERET. De plus, nous avons réalisé les tests sous forme de plusieurs protocoles expérimentaux. Chaque protocole correspond à un paramétrage particulier des attributs suivants :

- La base de données utilisée (AT&T ou Color FERET).
- La base de visages de référence (Galerie) : c'est l'ensemble des images de visage des personnes supposées connues par le système. Autrement dit, ce sont les vecteurs caractéristiques de ces images qui sont appris pendant la phase d'enrôlement.
- La base de visages de test : c'est l'ensemble des images de visage requêtes qui sont soumises à l'entrée du système pour la première fois pendant la phase d'identification.
- La méthode de classification adoptée pour les tests.
- Les résultats relevés.

Notons que pour ces différents protocoles expérimentaux, nous avons utilisé :

- Un banc de filtres de Gabor constitué de **40** filtres :
 - **8** orientations: $\theta_v = v\pi/8$ avec $v = \{0, 1, 2, \dots, 7\}$.
 - **5** fréquences centrales : $f_u = 0.25/2^{(u/2)}$ avec $u = \{0, 1, 2, \dots, 5\}$.
 - La dimension de la fenêtre de chaque filtre vaut: **128 x 128**.
- Un perceptron multicouche (PMC) ayant l'architecture et les paramètres suivants :
 - Une couche d'entrée dont le nombre de cellules d'entrées est égale à la taille des vecteurs caractéristiques qui dépend de la méthode d'extraction évaluée et la taille de la Galerie en question.
 - Une seule couche cachée composée d'un nombre de neurones qui est adapté aussi à la taille des vecteurs caractéristiques.
 - Une couche de sortie composée d'un nombre de neurones qui est égale à la taille de la Galerie.
 - La fonction **Sigmoïde** comme fonction d'activation des neurones.
 - Un critère d'arrêt de l'algorithme d'apprentissage qui correspond à une erreur quadratique inférieure à un seuil de 0.0001 ou un nombre d'itérations au maximum de 100000 itérations.
- Des SVM utilisant :
 - **$K.(K-1)/2$** classificateurs SVM binaires et non-linéaires (K est le nombre de classes (personnes) de la base de données référence (Galerie)).
 - La fonction noyau adoptée est de type Polynomial.
 - La méthode de classification multi-classes est un contre un.

Avant de décrire les détails de ces différents protocoles expérimentaux et leurs résultats, nous allons tout d'abord compléter la présentation des deux bases AT&T et Color FERET comme nous l'avons promis dans le chapitre 2.

7.3.1. La base AT&T

La base AT&T contient des images de visage (format PGM en niveau de gris) de 40 personnes, avec 10 images pour chacun (en total 400 images). Pour la plupart des sujets, les images ont été prises avec des variations différentes de l'apparence du visage : les expressions faciales (ouvert / fermé les yeux, souriant / pas souriant), les poses de tête et les détails du visage (lunettes / pas de lunettes). Les images prises n'ont pas en commun les mêmes types

de variation pour les différentes personnes de la base. En revanche, elles ont en commun la taille (92x 112) et le fond. Ces images sont organisées dans 40 répertoires (un pour chaque individu) nommé : s_1, \dots, s_{40} . Dans chacun de ces répertoires, il y a dix images de visage différentes pour chaque individu numérotées de 1 à 10. Un extrait de cette base est donné dans la figure 7.4.



Fig. 7.4. Extrait de la base AT&T

7.3.2. La base Color FERET

La base Color FERET contient en total 14 126 images de 1 199 personnes. C'est une extension de la base FERET. Les images ont été acquises dans des conditions différentes d'expression, de pose, d'éclairage et de temps. Ces images sont de taille 256 x 384 pixels, et les fichiers sont en format PPM (Portable Pixel Map). Les images de visage de face sont regroupées en 5 partitions et les images de visage de profil sont regroupées dans un ensemble nommé FERET Pose. Les différentes partitions de la base Color FERET que nous avons utilisées sont:

- La partition FA : Un ensemble de 994 images de visage, en vue frontale, prises lors de la première session FERET.
- La partition FB : Un ensemble contenant 992 images. Les sujets sont les mêmes que ceux de la partition FA sauf que les expressions faciales sont différentes. Les photos des partitions FA et FB ont été prises lors de la même session.
- La partition FC : Un ensemble contenant 194 images. Il contient les images de certaines personnes de la partition FA mais avec des variations en éclairage.
- L'ensemble FERET Pose contient des images de 200 personnes sous 9 poses différentes par personne ($-60^\circ, -40^\circ, -25^\circ, -15^\circ, 0, +15^\circ, +25^\circ, +40^\circ, +60^\circ$).

- Dans la nouvelle version de la base FERET (Color FERET), on trouve des images de visage (couleur) de quelques personnes qui sont prises avec 5 vues de pose différentes (-90° , -45° , 0° , $+45^\circ$, $+90^\circ$).

Des exemples extraits de la base Color FERET sont présentés dans la figure 7.5.

Pour utiliser les images de cette base, nous avons précédé à un recadrage des visages pour les centrer par rapport aux yeux de telle sorte que tous les visages aient leurs yeux approximativement au même endroit dans l'image et en même temps pour avoir des vignettes de visages de même dimension.



Fig. 7.5. Des exemples extraits de la base Color FERET

7.3.3. Protocole expérimental 1

Ce premier protocole expérimental vise deux objectifs. Le premier est l'évaluation des performances (taux de reconnaissance et rapidité) de la première approche *ACP-ADL double* et la deuxième approche *Gabor+ACP-ADL double* (sans estimateur de pose) en les comparant avec celle de la méthode classique *ACP (EigenFaces)*. Tandis que le deuxième objectif, consiste à établir une synthèse comparative des méthodes de classification déjà étudiées dans le chapitre 7, à savoir : les trois distances métriques (la distance city-block, la distance Euclidienne et la distance de Mahalanobis), les perceptrons multicouches (PMC) et les Séparateurs à Vastes Marges (SVM). Les attributs de ce protocole sont :

- La base de données utilisée : AT&T.
- La base de visages de référence (Galerie) : 200 images de visage de la base AT&T, les 5 premières images pour chaque personne (en total 5x40)
- La base de visages de test : 200 images de visage de la base AT&T, les 5 dernières images pour chaque personne (en total 5x40)
- Les méthodes de classification adoptées pour les tests : la distance city-block, la distance Euclidienne, la distance de Mahalanobis, PMC et SVM.

- Les résultats relevés :
 - Les courbes CMC de la méthode ACP en utilisant dans la classification l'une des trois distances citées en haut (figure 7.6).
 - Les courbes CMC de la méthode *ACP-ADL double* en utilisant dans la classification l'une des trois distances citées en haut (figure 7.7).
 - Les courbes CMC de la méthode *ACP-ADL double* en utilisant dans la classification PMC et SVM (figure 7.8).
 - Les courbes CMC de la méthode *Gabor+ACP-ADL double* en utilisant dans la classification l'une des trois distances citées en haut (figure 7.9).
 - Le taux de reconnaissance (taux de reconnaissance au rang 1) des trois méthodes **ACP**, **ACP-ADL double** et **Gabor+ACP-ADL double** avec les différentes méthodes de classification (Tableau 7.3).
 - Le temps moyen d'identification des trois méthodes **ACP**, **ACP-ADL double** et **Gabor+ACP-ADL double** avec les différentes méthodes de classification (Tableau 7.4). Ce temps est relevé entre l'instant de lecture de l'image de visage à identifier et l'instant de prise de décision.

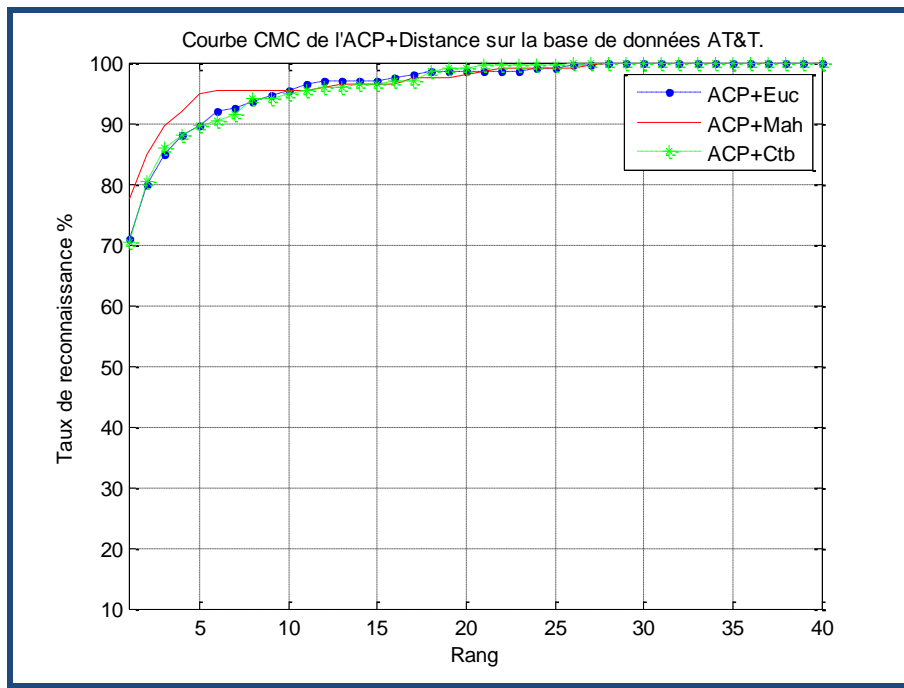


Fig. 7.6. Les courbes CMC de la méthode ACP en utilisant dans la classification l'une des distances : city-block, Euclidienne et Mahalanobis

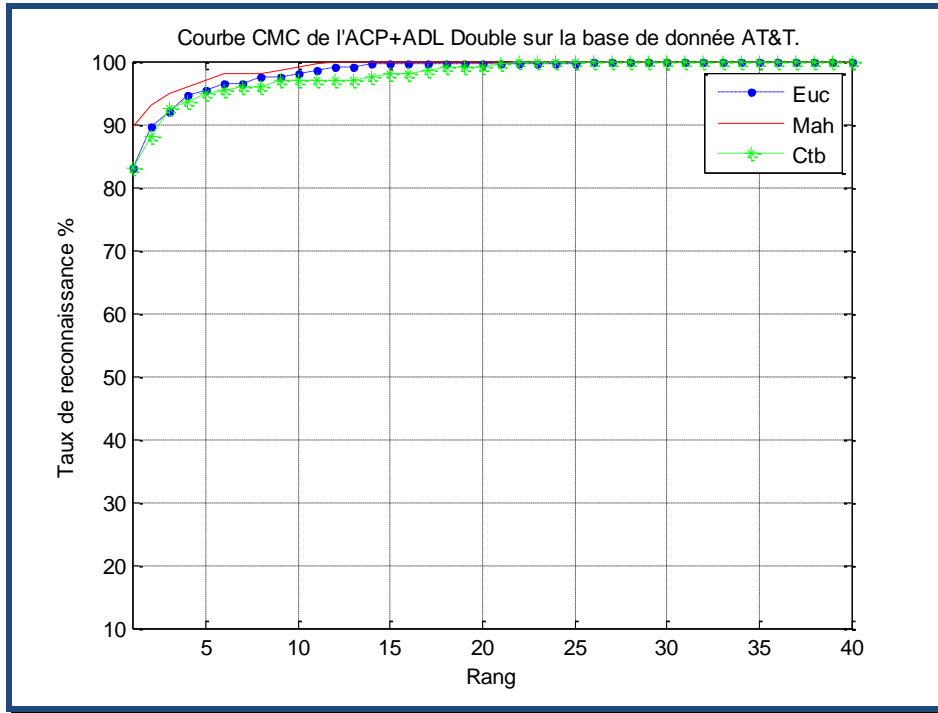


Fig. 7.7. Les courbes CMC de la méthode ACP-ADL double en utilisant dans la classification l'une des distances : city-block, Euclidienne et Mahalanobis

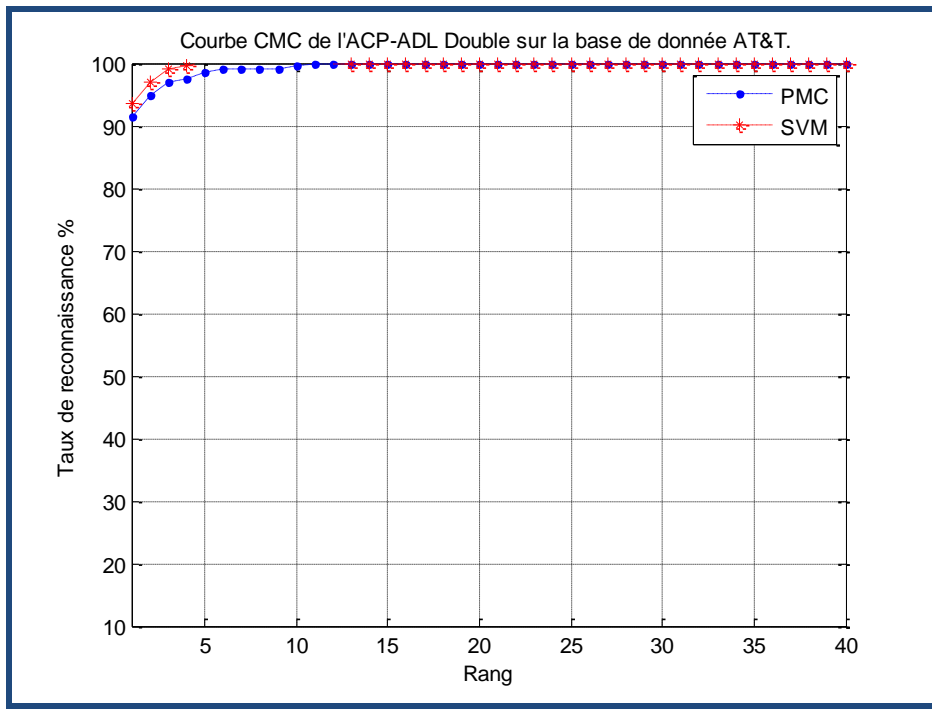


Fig. 7.8. Les courbes CMC de la méthode ACP-ADL double en utilisant dans la classification PMC ou SVM

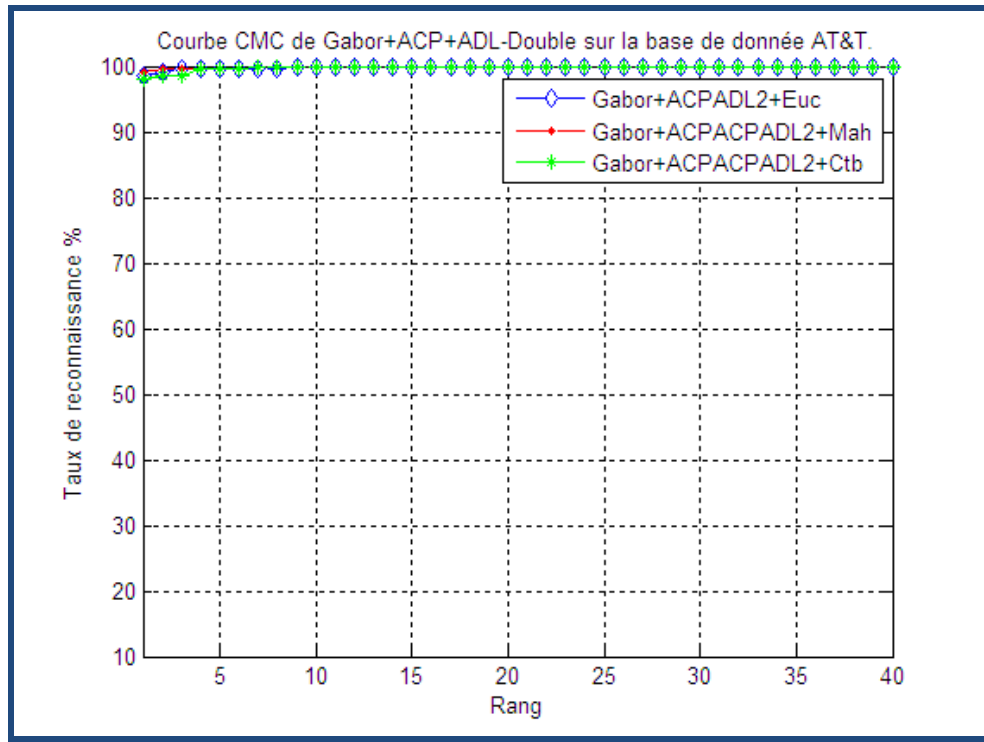


Fig. 7.9. Les courbes CMC de la méthode Gabor + ACP-ADL double en utilisant dans la classification l'une des distances : city-block, Euclidienne et Mahalanobis

Méthode d'extraction	Méthode de classification				
	CTB	Euc	Mah	PMC	SVM
ACP	70.50	71.00	77.50	79.00	80.50
ACP-ADL double	83.00	83.00	89.50	91.50	93.50
Gabor+ACP-ADL double	98.00	98.00	98.50	99.00	99.00

Tableau 7.3. Taux de reconnaissance des trois méthodes ACP, ACP-ADL double et Gabor + ACP-ADL double

Méthode d'extraction	Méthode de classification				
	CTB	Euc	Mah	PMC	SVM
ACP	0.0781	0.0782	0.0801	0.0810	0.0924
ACP-ADL double	0.0892	0.0890	0.0907	0.0911	0.1091
Gabor+ACP-ADL double	0.4974	0.4972	0.5003	0.5015	0.5186

Tableau 7.4. Temps moyen d'identification en (s) des trois méthodes ACP, ACP-ADL double et Gabor + ACP-ADL double

Pour faciliter l'interprétation des résultats du tableau 7.3, nous avons tracé l'histogramme de la figure 7.10.

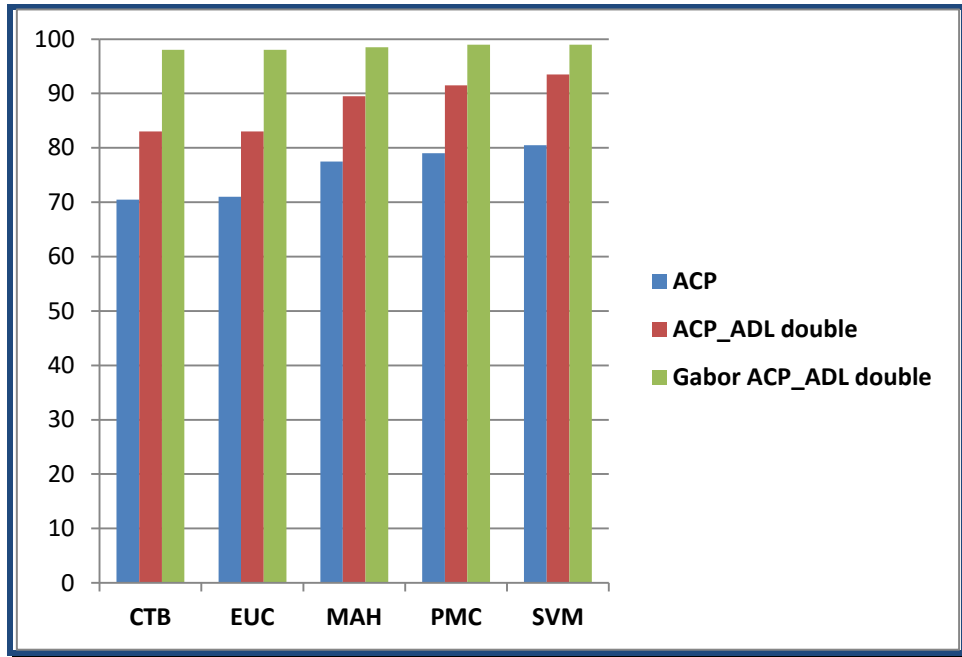


Fig. 7.10. Histogramme représentant les taux de reconnaissance des méthodes ACP, ACP-ADL double et Gabor + ACP-ADL double sur la base AT&T

Discussion

A partir de l'étude approfondie des résultats obtenus dans ce premier protocole, plusieurs constatations sont à relever que se soit pour les méthodes d'extraction ou pour les méthodes de classification.

Pour les méthodes d'extraction :

- Les taux de reconnaissances obtenus par la méthode *ACP-ADL double* surpassent ceux de la méthode ACP (la figure 7.10 et le tableau 7.3). La raison est que notre approche ne se contente pas uniquement de réduire l'espace de représentation des images de visages (comme le cas de l'ACP), mais elle assure en plus une bonne discrimination entre les classes de visages à l'aide des deux projections successives de l'ADL Fisher tout en minimisant les variations entre les visages de la même personne (intra-classe) et en maximisant les variations entre les visages des personnes différentes (inter-classes). Pour confirmer cette supériorité d'un autre angle, nous avons procédé aux projections des images de visage de 3 classes distinctes (3 personnes différentes avec 10 images par personne) successivement dans les trois sous-espaces de projection constituant l'*ACP-ADL double*. Les résultats de ces projections sont illustrés dans la figure 7.11. Toutefois, après un diagnostic détaillé des résultats obtenus,

nous avons remarqué que cette méthode n'arrive pas à identifier correctement tous les visages ayant des poses qui dépassent $\pm 25^\circ$ ou une forte variation d'illumination. L'explication de ce problème est donnée dans le paragraphe 5.2.2 du chapitre 5.

▪ Après la combinaison du **descripteur de Gabor** et **ACP-ADL double**, nous constatons qu'il y a une augmentation remarquable du taux de reconnaissance qui apparait de façon plus marquée sur l'histogramme de la figure 7.10 et le tableau 7.3. Donc, il est évident que le descripteur de Gabor a pu résoudre quelques problèmes de la méthode **ACP-ADL double** (par transitivité ceux de l'ACP). Cependant, Cette amélioration du taux de reconnaissance est au détriment du temps d'identification. Ce dernier comme nous le remarquons dans le tableau 7.4, augmente avec une valeur de l'ordre de **0.4s** ; c'est le temps nécessaire pour la convolution de l'image de visage d'entrée avec les **40** filtres de Gabor. En revanche, il y a une très légère augmentation de ce temps au niveau de l'ACP-ADL double par rapport à l'ACP.

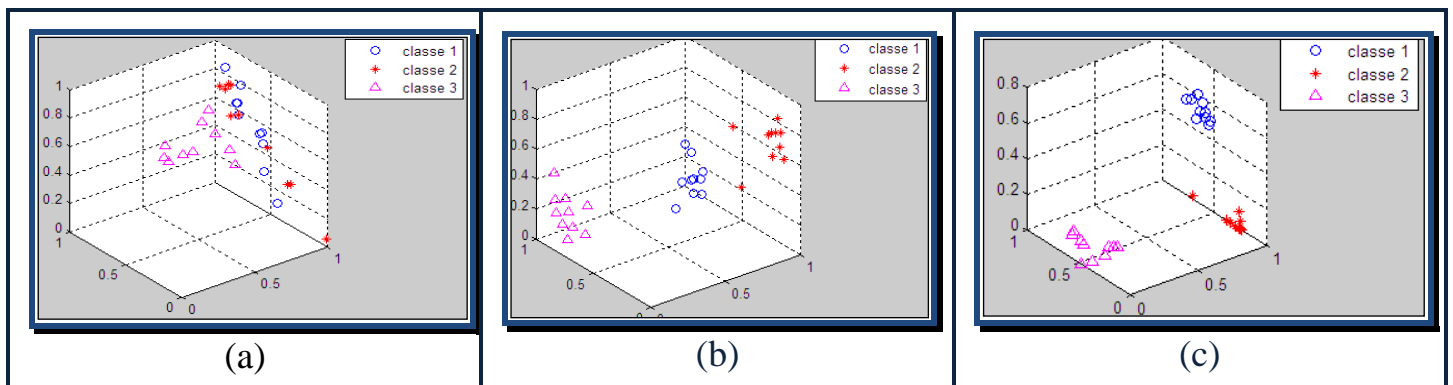


Fig. 7.11. Illustration des projections des images de visage de 3 personnes (10 images par personne ou classe) successivement dans les trois sous-espaces : a) le sous-espace de l'ACP, b) le sous-espace de l'ADL Fisher 1 et c) le sous-espace de l'ADL Fisher 2

Pour les méthodes de classification:

- Les résultats présentés dans le tableau 7.3 montrent que les deux méthodes de classification modernes (PMC et SVM) sont meilleures en termes du taux de classification que les méthodes classiques (les distances métriques). Par ailleurs, la distance de Mahalanobis se voit supérieure par rapport aux autres distances (la distance Euclidienne et la distance City-block). C'est pour cette raison que la majorité des travaux récents favorisent la distance de Mahalanobis par rapport aux autres distances.
- En ce qui concerne le temps de classification, nous constatons que les deux distances Euclidienne et City-block garantissent le temps le plus court. Contrairement aux SVM, ce temps est le plus long. Tandis que la distance de Mahalanobis et les PMC viennent au rang

milieu. Cette constatation ne peut pas être généralisée. En effet, lorsque la classification sera effectuée sur une base de données (Galerie) de taille très grande, ce temps de classification augmenterait considérablement pour les cas des distances métriques et les SVM car pour ces deux types de méthodes, le nombre de comparaisons croît avec le nombre de classes. En revanche, pour les PMC, ce temps est quasi-invariable puisque les PMC utilisent uniquement dans la phase d'identification des calculs de produit matriciel du vecteur caractéristique et les matrices représentant les poids des synapses. En conséquence, ce sont les PMC qui prendront le premier rang.

- Toutefois, les deux méthodes de classification : PMC et SVM, possèdent un inconvénient majeur qui se manifeste au moment de l'ajout d'une nouvelle personne dans la base de données référence (Galerie). Dans une telle situation, nous devons recommencer le processus d'apprentissage tout entier, à l'inverse des techniques qui utilisent une distance métrique, classifient et affectent des nouveaux visages sans passer par une étape d'apprentissage.

Les trois protocoles expérimentaux (2, 3, 4) qui vont être décrits juste après, ont en commun les attributs suivants :

- La base de données utilisée : FERET.
- La méthode de classification adoptée pour les tests : SVM.
- Le résultat relevé : taux de reconnaissance mentionné dans le tableau 7.5.

7.3.4. Protocole expérimental 2

Le but de ce protocole expérimental est d'évaluer le taux de reconnaissance des approches *ACP*, *ACP-ADL double* et *Gabor+ACP-ADL double* face à la variation d'expression faciale.

- La base de visages de référence (Galerie) : 100 images de visage de la base FERET de 50 personnes (une image de la partition FA et une image de la partition FC pour chaque personne).
- La base de visages de test : 50 images de visage de la base FERET des mêmes 50 personnes de la Galerie (une image de la partition FB pour chaque personne).

7.3.5. Protocole expérimental 3

L'objectif de ce protocole expérimental est d'évaluer la robustesse des approches *ACP*, *ACP-ADL double* et *Gabor+ACP-ADL double* face à la variation d'illumination.

- La base de visages de référence (Galerie) : 100 images de visage de la base FERET de 50 personnes de la Galerie (une image de la partition FA et une image de la partition FB pour chaque personne).
- La base de visages de test : 50 images de visage de la base FERET des mêmes 50 personnes (une image de la partition FC pour chaque personne).

7.3.6. Protocole expérimental 4

Dans ce protocole expérimental, nous allons évaluer la fiabilité des approches *ACP*, *ACP-ADL double* et *Gabor+ACP-ADL double* (sans estimateur de pose) face à la variation de pose.

- La base de visages de référence (Galerie) : 60 images de visage de la base FERET de 30 personnes (une image de la partition FA et une image de la partition FB pour chaque personne).
- La base de visages de test : 270 images de visage de l'ensemble FERET Pose de 30 personnes (9 image par personne).

Méthode d'extraction	Protocole		
	Variation d'expression faciale Protocole 2	Variation d'illumination Protocole 3	Variation de pose Protocole 4
ACP	64,50	51,00	33,40
ACP-ADL double	82,00	66,50	56,60
Gabor+ ACP-ADL double	90,00	87,50	81,30

Tableau 7.5. Taux de reconnaissance des méthodes ACP, ACP-ADL double et Gabor + ACP-ADL double avec différentes variations de l'apparence du visage.

Pour faciliter l'interprétation des résultats du tableau 7.5, nous avons tracé l'histogramme de la figure 7.11.

Discussion

En analysant les résultats présentés dans le tableau 7.5 et l'histogramme de la figure 7.11, qui récapitulent les taux de reconnaissance relevés au niveau des trois derniers protocoles (2, 3 et 4), nous constatons qu'il y a une grande amélioration du taux de reconnaissance avec l'approche *Gabor+ACP-ADL double* même en présence des fortes variations de l'apparence du visage de type : expression faciale, illumination et de pose. Ceci prouve de plus l'intérêt de

la combinaison du **descripteur de Gabor** et l'*ACP-ADL double* dans l'approche d'extraction des caractéristiques.

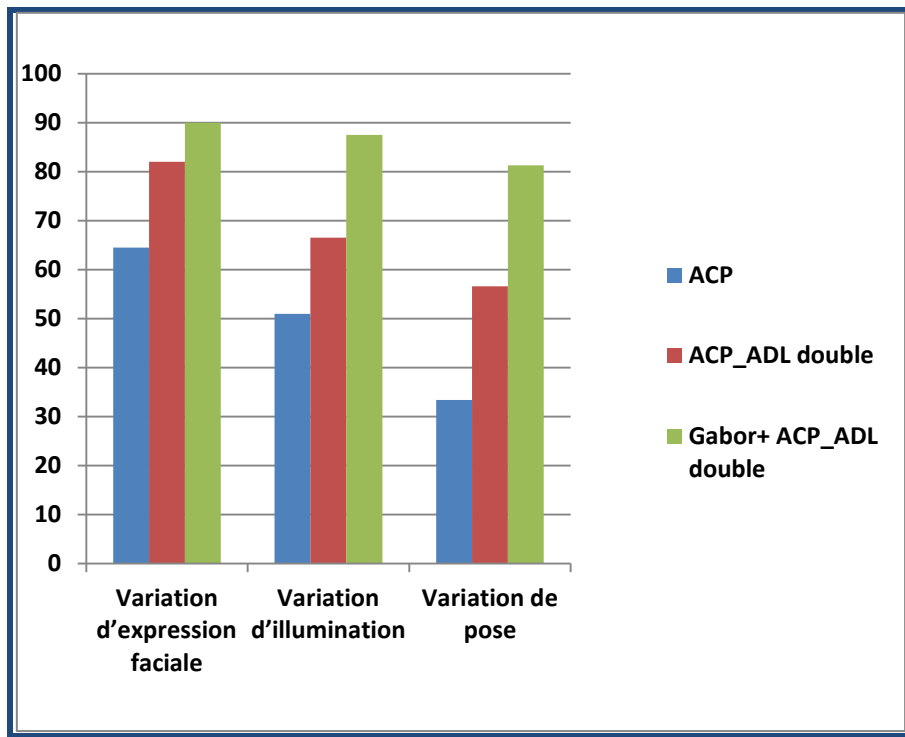


Fig. 7.12. Histogramme représentant les taux de reconnaissance des méthodes ACP, ACP-ADL double et Gabor + ACP-ADL double sur la base Color FERET

Le taux de reconnaissance de **90%** obtenu dans le protocole expérimental 2 (variation d'expression faciale) est un taux très satisfaisant.

Le taux de reconnaissance de **87,50%** atteint dans le protocole expérimental 3, reste un taux raisonnable en tenant compte des fortes variations d'illumination présentes au niveau des images de la partition FC de la base Color FERET (base de Test). Néanmoins, nous pouvons améliorer ce taux si nous procédons à un prétraitement des images avant l'extraction des caractéristiques. A titre d'exemple, nous avons soumis un ensemble d'images de test à un égalisateur d'histogramme (voir paragraphe 4.8.2 du chapitre 4). Les visages de cet ensemble ne sont pas correctement identifiés avant le prétraitement. An contraire, après ce dernier, nous constatons qu'une partie de cet ensemble le devienne.

Le taux de reconnaissance de **81,30 %** obtenu avec notre approche **Gabor+ACP-ADL double** (sans estimateur de pose) dans le protocole expérimental 4 (variation de pose) est un taux modeste. La cause est que l'approche n'arrive pas à identifier correctement la plupart des visages ayant une pose de 60°. Une solution raisonnable pour remédier ce problème, consiste

dans l'intégration de l'estimateur de pose proposé, et déjà évoqué comme un outil primordial, dans l'architecture de la deuxième approche d'extraction détaillée dans la section 5.3 du chapitre 5.

Les tests expérimentaux concernant cet estimateur de pose sont réalisés lors du dernier protocole expérimental 5 présenté ci-après.

7.3.7. Protocole expérimental 5

Nous clôturons la présentation des expérimentations réalisées dans le cadre cette thèse par un protocole qui va permettre d'évaluer les performances de l'estimateur de pose proposé. Par conséquent, valider son impact sur la chaîne d'identification, en particulier dans un environnement où il ya des fortes variations de pose.

La structure de l'estimateur de pose proposé dans la section 5.3.4 du chapitre 5, est conçue pour estimer d'une façon discrète la pose d'un visage et donner la valeur la plus proche parmi les 7 valeurs suivantes: -30° , -60° , -90° , 0° , $+30^\circ$, $+60^\circ$, $+90^\circ$. Notons que dans les tests, nous avons adopté les valeurs: -25° , -60° , -90° , 0° , $+25^\circ$, $+60^\circ$, $+90^\circ$. Ce choix est imposé par les grandeurs de pose présentes dans la base Color FERET.

Pour assurer l'apprentissage des classificateurs SVM de l'estimateur, nous avons collecté un ensemble de 70 images de visage (les poses modèles) pour former 7 groupes. Chaque groupe contient 10 images de visage qui ont des orientations les plus proches pour définir finalement 7 classes $\{-25^\circ, -60^\circ, -90^\circ, 0^\circ, +25^\circ, +60^\circ, +90^\circ\}$.

Lors des tests, nous avons soumis à l'entrée de l'estimateur de pose une série de 55 images de visage extraites aléatoirement de la base Color FERET (les poses de test). Les visages contenus dans ces images ont des poses distinctes: -90° , -60° , -45° , -25° , -15° , 0° , $+15^\circ$, $+25^\circ$, $+45^\circ$, $+60^\circ$, $+90^\circ$ (10 images par pose). Du point de vue théorique, pour ces valeurs de pose de test, nous devons aboutir à la classification suivante:

Valeur de la pose de test	$-90^\circ/+90^\circ$	$-60^\circ/+60^\circ$ $-45^\circ/+45^\circ$	$-25^\circ/+25^\circ$ $-15^\circ/+15^\circ$	0°
Sortie de l'estimateur	$-90^\circ/+90^\circ$	$-60^\circ/+60^\circ$	$-25^\circ/+25^\circ$	0°

Tableau 7.6. La classification théorique des valeurs de pose de test

Les résultats des classifications expérimentales obtenus par l'estimateur de pose sont donnés par les deux taux suivants ;

- Le taux des correctes estimations : **92,72%**
- Le taux des fausses estimations : **7.27%**

Discussion

Lors des tests de ce dernier protocole, nous avons soulevé que les fausses estimations de pose concernent **4** images dont la pose vaut **-45°** ou **+45°**. Cela est dû probablement à l'emplacement de cette pose dans une zone de chevauchement des deux classes : **-60°** et **-25°** ou **+25°** et **+60**. Toutefois, malgré ce problème, nous pouvons déclarer l'efficacité notable de notre estimateur de pose développé étant donné la valeur du taux des estimations correctes obtenu.

7.4. Conclusion

Dans ce chapitre nous avons effectué des expérimentations détaillées avec les approches qui ont été proposées dans ce mémoire. Ces expérimentations ont été menées sur des bases de données standards, à savoir: Caltech_10K_WebFaces, AT&T et Color FERET. Par la suite, nous avons évalué les performances de ces approches en les comparant à celles obtenues par d'autres méthodes de l'état de l'art grâce aux différents protocoles expérimentaux mises en œuvre. En conséquence, nous avons pu établir certaines conclusions intéressantes à la fin de ces expérimentations. Nous dresserons, dans le chapitre suivant, le bilan de ces conclusions et nous exposerons quelques perspectives.

Bilan et Perspectives

Conclusion générale

Cette thèse a abordé trois domaines clés de la reconnaissance de visages dans des images 2D, ceux de la détection de visage, l'extraction des vecteurs caractéristiques et la classification. Tout au long de cette thèse, nous avons développé une étude de l'existant dans ces domaines. Nous avons proposé par la suite des solutions aux problèmes touchant aux deux premiers et une synthèse pour le dernier, dans l'objectif de développer un système automatique de reconnaissance de visages (partie logicielle) plus satisfaisant en terme de fiabilité et de rapidité.

Nous avons proposé dans un premier temps, une approche pour la détection et la localisation des visages dans des images 2D à multiple visages. A partir de l'image couleur captée, nous construisons une image binaire composée de deux classes de pixels (peau et non-peau) à l'aide d'un algorithme de détection des pixels de peau. Dans cet algorithme, nous avons mis en œuvre un nouveau modèle de représentation de la distribution de la couleur de peau. Ensuite, une segmentation en régions de peau connexes et homogènes est effectuée sur l'image binaire en utilisant un célèbre et efficace algorithme en segmentation, la Ligne de Partage des Eaux (*LPE*). Ce dernier génère un ensemble de régions de peau connexes dont nous masquons et nous sélectionnons uniquement les plus significatives. Les imagerie englobant ces régions sont localisées et soumises à une procédure de classification visage et non-visage. Dans cette procédure, nous avons opté pour la solution du « Template-matching » qui consiste dans la détermination de la correspondance entre l'objet (visage ou non-visage) figurant dans l'imagerie et le modèle de visage que nous avons élaboré. La définition de ce modèle est inspirée d'une part de l'apparence globale d'un visage type (la présence des yeux et la bouche) et d'autre part des règles géométriques appliquées sur les positions relatives de ces traits. La vérification de la présence de ces traits et leur localisation sont assurées par une cascade de filtres de Gabor.

Les résultats obtenus affichent une validation de la démarche scientifique suivie pour intégrer tous ces outils dans une telle approche. Nous avons réussi à réaliser un taux de détection très satisfaisant avec un temps de calcul minimal par rapport au célèbre algorithme de détection de visage, celui de Viola et Jones. De plus, cette approche ne nécessite ni base de données, ni phase d'apprentissage. Toutefois, elle se limite dans la détection des visages dont

l'orientation ne dépasse pas $\pm 60^\circ$. La cause de ce handicap est expliquée dans le chapitre précédent.

Nous nous sommes également intéressés au développement d'une approche d'extraction des vecteurs caractéristiques (signatures) des images de visage. La première version de cette approche porte le nom d'*ACP-ADL double* ; l'ACP (Analyse en Composantes Principales ou '*EigenFaces*') combinée avec l'ADL de Fisher (Analyse Discriminante Linéaire de Fischer) doublée deux fois successives. Cette approche est basée sur la définition de trois nouveaux sous-espaces (sous-espace des faces propres et deux sous-espaces de Fisher) dans lesquelles nous procédons à trois projections successives des images de visage. L'importance de cette combinaison réside dans le point fort de l'ACP et l'ADL qui est la réduction de dimensionnalité de l'espace de représentation de l'image de visage et la puissance de l'ADL dans la discrimination entre les classes de visage après projection (réduction). Cette méthode a abouti à des performances très encourageantes en termes de rapidité dans la phase de reconnaissance et du taux de reconnaissance lorsqu'elle est testée sur une base de données des images de visages qui ne possèdent pas des fortes variations de pose ou d'illumination. Toutefois, ce taux chute rapidement dans le cas contraire.

Nous nous sommes donc tournés vers une autre solution pour surmonter ce problème. Elle consiste à conserver la première version comme moyen de réduction de dimensionnalité et introduire deux améliorations intéressantes. La première est l'introduction du descripteur de Gabor. Chaque image de visage de la base de données ou de test, pour qu'elle soit modélisée avec ce descripteur, subit une convolution avec un banc de filtres de Gabor. Les réponses de cette opération de convolution vont constituer cette fois-ci les données de départ pour construire les sous-espaces de réduction *ACP-ADL double*. Tandis que la deuxième amélioration proposée est effectuée en amont de la phase de classification par l'intégration d'un estimateur de pose pour s'affranchir du problème des fortes variations de la pose.

Le but de cette solution est d'adapter la base de données (Galerie) en fonction de la pose du visage à reconnaître. Il s'agit de disposer d'une multitude de galeries paramétrées par la pose des visages et chacune possédant son propre classificateur. Dans cette nouvelle situation, l'estimateur de pose a pour rôle de cibler le classificateur associé à la plage de pose qui contient la pose du visage à reconnaître. Notre estimateur de pose développé pour cette finalité s'est basé sur la construction d'un classificateur par un apprentissage supervisé en utilisant les séparateurs à vastes marges (SVM). Cet apprentissage a eu lieu à l'aide des

exemples de visage pour différentes pose modélisées par un descripteur de pose fondé sur quelques ondelettes particulières de Gabor qui portent l'information de pose.

Les résultats que nous avons présentés dans le chapitre précédent montrent effectivement l'intérêt de ces deux améliorations qui ont permis d'aboutir finalement à une nouvelle approche d'extraction des vecteurs caractéristiques plus robuste face aux variations de pose, d'illumination et d'expression. Cette robustesse se manifeste pratiquement dans les précisions apportées en terme de taux de reconnaissance et se justifie théoriquement par :

- La pertinence de la modélisation d'un visage par le descripteur de Gabor qui minimise les effets liés aux variations d'illumination, d'expression et de pose qui ne dépasse pas $\pm 40^\circ$.
- L'utilisation de l'information de pose comme paramètre d'entrée dans la phase de classification des vecteurs caractéristiques facilite la reconnaissance de visages pris dans des poses variables.
- La grande puissance de la première version d'extraction (*ACP-ADL double*) dans la réduction de la dimension des vecteurs extraits par le descripteur de Gabor, sans perte d'information et avec une discrimination inter-classes.

En ce qui concerne la classification, la synthèse élaborée sur les méthodes de classification : les distances métriques, les réseaux de neurones et les séparateurs à vastes marges (SVM), nous a permis de déduire l'importance des deux derniers dans la phase de classification. Les expérimentations qui ont été menées sur ces méthodes, affirment que les SVM garantissent la meilleure efficacité de classification (un grand taux de classification) par rapport aux réseaux de neurones. En revanche, ce sont ces derniers qui garantissent la meilleure rapidité (un très court temps de calcul dans la phase de reconnaissance). Cette comparaison nous donne un point de départ pour faire un choix optimal qui va dépendre de l'application envisagée.

En conclusion, les études et les contributions faites dans les trois domaines cités en haut, nous ont permis d'acquérir une compréhension plus approfondie du sujet traité et d'enrichir nos connaissances au niveau de plusieurs disciplines : biométrie, traitement d'image, intelligence artificielle, analyse de données, morphologie mathématique,

Perspectives

Dans les perspectives de notre travail, plusieurs axes de recherche nous semblent prometteurs. De point de vue théorique, cette thèse ouvre plusieurs pistes pour des futurs

travaux de recherche. Nous proposons pour le moment deux améliorations qui peuvent être apportées à notre travail :

La première amélioration consiste dans le traitement de l'étape d'estimation de pose dans le module de la détection des visages au lieu de la traiter en amont de la classification. Nous proposons aussi de combiner la procédure de classification de visage et non-visage avec celle de l'estimation de pose dans un seul processus. Ceci peut être réalisé à l'aide d'un Boosting de classificateurs SVM. On peut remplacer la cascade des filtres de Gabor par une cascade de classificateurs SVM. Ses premiers étages auront comme tâche la classification des objets figurants dans les régions de peau segmentées, en deux classes : visage et non-visage. Si ces étages nous confirment que c'est un visage, on poursuivra le traitement avec les étages postérieurs. Ces derniers auront comme objectif l'estimation de la pose de ce visage suite à une classification. L'apprentissage de ces classificateurs se fera avec des descripteurs qui seront capables de discriminer en même temps la nature d'un objet (visage et non-visage) et la pose d'un visage. Avec cette amélioration, on peut garantir un gain en temps de calcul et une détection des visages pour n'importe quelle pose.

La deuxième amélioration qui nous semble abordable peut être menée dans le module d'extraction. Le descripteur de Gabor a montré dans notre approche d'extraction des caractéristiques son efficacité dans la tolérance de certains types de variations (l'illumination, l'expression et la pose) à un certain ordre. Or, il serait intéressant de fusionner ce descripteur avec un autre en utilisant la stratégie de fusion avec pondération de score. Ce nouveau descripteur doit avoir la capacité de tolérer d'autres types de variation d'apparence d'un visage. Avec cette amélioration on peut réussir à développer un algorithme d'extraction des caractéristiques robuste face à un maximum d'artéfacts. Bien sûr, ce travail doit être réalisé dans le sens où la rapidité du système de reconnaissance ne se dégrade pas.

Du point de vue pratique, nous suggérons une extension de notre système de reconnaissance de visage développé aux séquences vidéo. En effet, si on veut envisager notre système conçu pour une application d'identification telle que la vidéosurveillance, il serait obligatoire de le tester sur des séquences de vidéo prises dans un environnement réel non contrôlé. Dans ce contexte, on aura l'occasion de traiter toutes les artéfacts possibles en même temps, ainsi que l'aspect temporel de la vidéo.

Annexe A

Espaces de couleurs

Pour modéliser la distribution de la couleur de peau, plusieurs espaces de couleurs populaires ont été proposés dans la littérature. Nous présentons dans cette annexe les espaces les plus utilisés et que nous avons comparés expérimentalement pour valider les espaces adoptés dans notre détecteur des pixels de peau.

A.1. L'espace RVB

L'espace RVB a été mis au point en 1931 par la Commission Internationale de l'Eclairage (CIE). Il consiste à représenter l'espace des couleurs à partir de trois composantes Rouge-Vert-Bleu. Cet espace correspond à la façon dont les couleurs sont généralement codées informatiquement, ou plus exactement à la manière dont les écrans à tubes cathodiques (ordinateurs, TV) représentent ces couleurs. C'est l'espace de couleur standard pour les images numériques. En effet, chaque pixel dans une image numérique est codé par trois nombres décimaux compris entre **0** et **255** qui décrivent quelles quantités de rouge, de vert et de bleu (**R**, **G**, **B** $\in [0,255]$) le composent. Nous pouvons, grâce à ces trois composantes primaires, définir toute couleur **C** comme la somme des quantités de chaque couleur primaire. La représentation des couleurs dans cet espace donne un cube appelé cube de Maxwell (figure A.1).

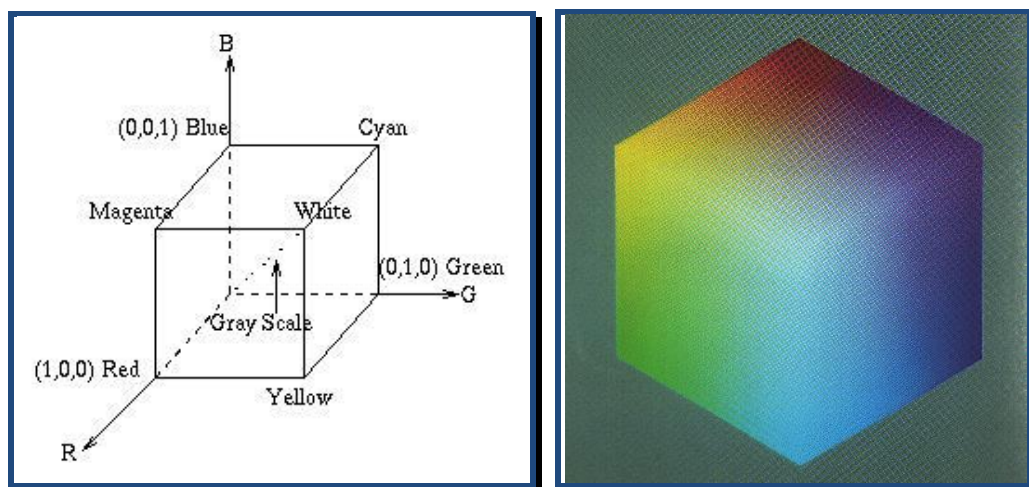


Fig. A.1. Cube de Maxwell

A.2. L'espace RGB normalisé

L'espace RGB normalisé (\mathbf{r} , \mathbf{g} , \mathbf{b}) est défini comme ceci :

$$\begin{cases} \mathbf{r} = \frac{R}{R+G+B} \\ \mathbf{g} = \frac{G}{R+G+B} \\ \mathbf{b} = \frac{B}{R+G+B} \end{cases} \quad (\text{A.1})$$

A.3. L'espace HSV

Le principe de l'espace **HSV** (*Hue*, *Saturation*, *Value*) est de caractériser les couleurs de façon plus intuitive, conformément à la perception naturelle des couleurs par l'œil humain et comment les artistes mélangent les couleurs typiquement. Le modèle HSV consiste à décomposer la couleur selon des critères physiologiques:

- La teinte (en Anglais Hue) : correspondant à la perception de la couleur. Intuitivement, c'est le nom qu'on utilisera pour désigner la couleur, "vert", "mauve", "orange", etc.
- La saturation (en Anglais Saturation): c'est le taux de pureté de la couleur, c'est-à-dire son caractère vif ou terne, qui doit varier entre la pureté maximale (couleur éclatante) et l'achromatisme (niveau de gris).
- La valeur (en Anglais Value): c'est la mesure de l'intensité lumineuse de la couleur, qui doit varier entre le noir absolu et le blanc.

Le passage de l'espace RGB à l'espace HSV est exprimé par les relations suivantes :

$$\begin{cases} H = \arccos \frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{((R-G)^2+(R-G)(G-B))}} \\ S = 1 - 3 \frac{\min(R,G,B)}{R+G+B} \\ V = \frac{1}{3}(R + G + B) \end{cases} \quad (\text{A.2})$$

A.3. L'espace YCrCb

Dans cet espace la luminance (c'est-à-dire l'information en noir et blanc) présentée par la composante Y, est séparée de la couleur, tandis que Cr et Cb permettent de représenter la chrominance, c'est-à-dire l'information couleur. Cet espace est souvent utilisé dans la compression des images. Les paramètres de l'espace YCrCb sont codés à partir des composantes de l'espace RVB en utilisant la conversion suivante :

$$\begin{cases} Y = 0.299 R + 0.587 G + 0.114 B \\ Cb = -0.169 R - 0.331 G + 0.500 B \\ Cr = 0.500 R - 0.419 G - 0.081 B \end{cases} \quad (\text{A.3})$$

Annexe B

Algorithme de rétro-propagation du gradient

B. 1. Principe de l'algorithme de rétro-propagation du gradient

La rétro-propagation du gradient consiste à propager à l'envers (de la couche de sortie vers la couche d'entrée) l'erreur obtenue sur les exemples de la base d'apprentissage. On utilise pour cela l'erreur quadratique (le carré de la différence entre ce qu'on obtient et ce qu'on désire). Si on calcule la dérivée partielle de l'erreur quadratique par rapport aux poids des connexions, il sera possible de déterminer la contribution des poids à l'erreur générale, et de corriger ces poids de manière à se rapprocher du résultat souhaité. La correction se fait par itération en corrigeant plus ou moins fortement les poids par l'intermédiaire d'un coefficient. A l'issue d'un certain nombre d'itérations, lorsque qu'on est satisfait du classement des exemples de notre base d'apprentissage, on fixe les poids constituant ainsi des frontières inter-classes.

B.2. Etapes de l'algorithme

Pour expliquer les étapes de cet algorithme, considérons le perceptron multicouche à une couche cachée représenté par la figure B1. Ce perceptron possède :

- Une couche d'entrée à n cellules d'entrées X_i (Il ne s'agit pas de neurones, ces cellules présentent simplement les entrées du perceptron).
- Une couche cachée composée de m neurones à m sorties Y_j .
- Une couche de sortie composée de p neurones à p sorties Z_k .
- $n \times m$ connexions entre la couche d'entrée et la couche cachée, chacune pondérée par un poids synaptique V_{ij} .
- $m \times p$ connexions entre la couche cachée et la couche de sortie, chacune pondérée par un poids synaptiques W_{jk}

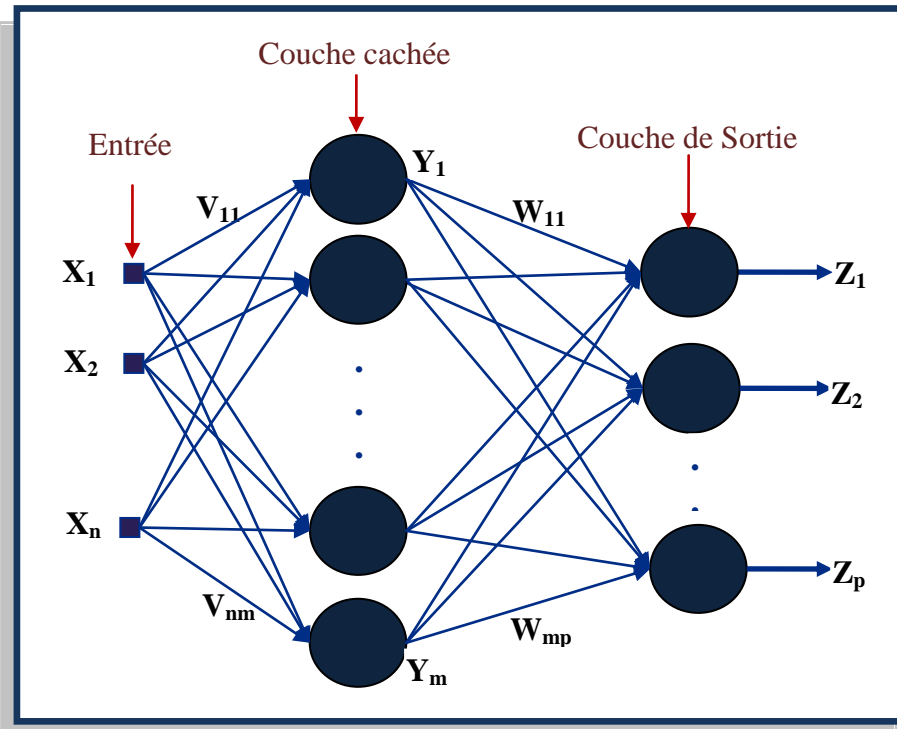


Fig. B. 1. Perceptron multicouche à une seule couche cachée : n cellules d'entrée, m neurones dans la couche cachée et p neurones dans la couche de sortie

Etape 1 : Initialisation des poids des connexions

Les poids synaptiques V_{ij} et W_{jk} sont initialisés au hasard.

Etape 2 : Propagation des entrées

Chaque exemple de la base d'apprentissage est présenté à la couche d'entrée. On le propage vers la couche cachée :

$$Y_j = f\left(\sum_{i=1}^n X_i V_{ij} + X_0\right) \quad (\text{B.1})$$

Puis de la couche cachée vers la couche de sortie :

$$Z_k = f\left(\sum_{j=1}^m Y_j W_{jk} + Y_0\right) \quad (\text{B.2})$$

Les valeurs X_0 et Y_0 sont des biais : des scalaires et non des sorties de la couches précédente.

f est la fonction d'activation choisie.

Etape 3 : Rétro-propagation de l'erreur

Pour chaque exemple de la base d'apprentissage appliqué en entrée du réseau, on calcule son erreur au niveau des couches de sortie, c'est à dire la différence entre la sortie désirée S_k et la sortie réelle Z_k :

$$E_k = Z_k(1 - Z_k)(S_k - Z_k) \quad (\text{B.3})$$

On propage cette erreur sur la couche cachée ; l'erreur de chaque neurone de la couche cachée est donnée par :

$$F_j = F_j(1 - Y_j) \sum_{k=1}^p W_{jk} \cdot E_k \quad (\text{B.4})$$

Etape 4 : Correction des poids des connexions

Il reste à modifier les poids des connexions :

Entre la couche d'entrée et la couche cachée :

$$\Delta W_{jk} = \eta \cdot Y_j \cdot E_k \quad \text{et} \quad \Delta Y_0 = \eta \cdot E_k \quad (\text{B.5})$$

Entre la couche cachée et la couche de sortie :

$$\Delta V_{ij} = \eta \cdot X_i \cdot F_j \quad \text{et} \quad \Delta Y_0 = \eta \cdot F_j \quad (\text{B.6})$$

η étant un paramètre qu'il faut déterminer empiriquement.

Boucler à l'étape 2 jusqu'à un critère d'arrêt à définir.

Le critère d'arrêt n'est pas forcément évident à définir. Comme exemple, on décide d'arrêter quand les erreurs résiduelles sont inférieures à un certain seuil ou le nombre d'itérations atteint une valeur prédéfinie en avance.

Diffusion scientifique de la recherche

Nous soulignons dans cette section le résultat de la diffusion des travaux de cette recherche dans des médiums scientifiques.

Articles de revues internationales

A. HAJRAOUI, M. SABRI et M. FAKIR. **Complete architecture of a robust system of face recognition**. International Journal of Computer Applications. Vol. 122, N°. 1, pages: 26-31, Juin 2015.

A. HAJRAOUI et M. SABRI. **Face Detection Algorithm based on Skin Detection, Watershed Method and Gabor Filters**. International Journal of Computer Applications. Vol. 94, N°. 6, pages: 33-39, Mai 2014.

A. HAJRAOUI, M. SABRI, M. FAKIR et O. BENCHAREF. **A new approach for Face Recognition Based on PCA & Double LDA Treatment combined with SVM**. IOSR Journal of Engineering. Vol. 2, N°.4, pages: 685-691, Avril 2012.

Communications orales de conférences internationales

A. HAJRAOUI, M. SABRI, M. FAKIR, B. BOUIKHALENE et S.SAFI. **Approche ACP.ADL et réseaux de neurones appliquée à l'analyse des visages**. TELECO2011 & 7ème JFMMA, Mars 16-18, 2011 – Tanger MAROC.

Bibliographie

- [1] A. Ceyhan et P. Piazza. L'identification biométrique : Champs, acteurs, enjeux et controverses. ISBN: 9782735113156. Editeur: Éditions de la Maison des sciences de l'homme, 2011.
- [2] A. K. Jain, A. A. Ross et K. Nandakumar. Introduction to Biometrics. e-ISBN: 978-0-387-77326-1. Editeur: Springer, 2011.
- [3] J. Pierson. La biométrie : L'identification par le corps. ISBN-13: 978-2746218314. Editeur: Hermes Science Publications. Septembre 2007.
- [4] A. K. Jain, A. Ross et S. Prabhakar. An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, Vol. 14, No. 1, Janvier 2004.
- [5] J. Hennebert. Biométrie : Modalités et Applications. Conférence du Groupement Industriel Fribourgeois, Décembre 2008.
- [6] P. J. Phillips, P. Rauss et S. Der. FERET (Face Recognition Technology) Recognition Algorithm Development and Test Report. ARL-TR-995, U.S. Army Research Laboratory, Octobre 1996.
- [7] P. J. Phillips, H. Wechsler, J. Huang et P. J. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. Image Vision Computer, Vol. 16, N°. 5, pages: 295-306, 1998.
- [8] D. M. Blackburn, J. M. Bone et P. J. Phillips. Face Recognition Vendor Test 2000. FRVT 2000 Evaluation Report, NIST, Février 2001.
- [9] M. Ngan et P. Grother. Face Recognition Vendor Test (FRVT), Performance of Automated Age Estimation Algorithms. NIST Interagency Report. Mars, 2014.
- [10] E. Barbeau, S. Joubert et O. Felician. Traitement et reconnaissance des visages : du percept à la personne. SOLAL, janvier 2009.
- [11] J. Y. Baudouin, V. Chambon et G. Tiberghien. Expert en visages ? Pourquoi sommes-nous tous des experts en reconnaissance des visages. L'évolution Psychiatrique, Vol. 74, Issue 1, pages : 3–25, Janvier –Mars 2009.
- [12] X. Ding et C. Fang. Discussions on Some Problems in Face Recognition. Advances in Biometric Person Authentication, Lecture Notes in Computer Science, Vol. 3338, pages: 47-56, Springer Berlin Heidelberg, 2005.
- [13] M. Hassaballah et S. Aly. Face Recognition: Challenges, Achievements, and Future Directions. IET Computer Vision, pages: 1-15, Mai 2015.

-
- [14] X. Zhang et Y. Gao. Face recognition across pose: A review. *Pattern Recognition*, Vol. 42, pages: 2876-2896, Novembre 2009.
- [15] X. Tan et B. Triggs. Enhanced Local Texture Feature Sets for Face Recognition under difficult Lighting Conditions. Dans *Analysis and Modelling of Faces and Gestures*, tome 4778 de LNCS, pages: 168–182. Springer, Octobre 2007.
- [16] N. S. Vu et A. Caplier. Illumination-robust face recognition using retina modeling. Dans *16th IEEE International Conference on Image Processing (ICIP 2009)*, pages 3289–3292, Novembre 2009.
- [17] P. Santemiz, L. J. Spreeuwers et N. J. R. Veldhuis. Side-view face recognition. *Proceedings of 32nd WIC Symposium on Information Theory in the Benelux*, 10–11 Mai 2011.
- [18] S. J. D. Prince, J. H. Elder, J. Warrell et F. M. Felisberti: Tied Factor Analysis for Face Recognition across Large Pose Differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, pages: 970–984, 2008.
- [19] A. Li, S. Shan, X. Chen et W. Gao. Cross-pose face recognition based on partial least squares. *Pattern Recognition Letter*, Vol. 32, Issue 15, pages: 1948–1955, 2011.
- [20] A. Sharma, M. Al Haj, J. Choi, L. S. Davis et D. W. Jacobs. Robust pose invariant face recognition using coupled latent space discriminant analysis. *Computer Vision and Image Understanding*, Vol. 116, pages: 1095–1110, 2012.
- [21] K. Hotta. Robust face recognition under partial occlusion based on support vector machine with local Gaussian summation kernel. *Image and Vision Computing*, Vol. 26, pages: 1490-1498, Novembre 2008.
- [22] E. Norouzi, M. N. Ahmadabadi et B. N. Araabi. Attention control with reinforcement learning for face recognition under partial occlusion. *Machine Vision and Applications*, Vol. 22, Issue 2, pages: 337-348, Mars 2011.
- [23] J. Shermina et V. Vasudevan. Face Recognition System with Various Expression and Occlusion Based on a Novel Block Matching Algorithm and PCA. *International Journal of Computer Applications*, Vol. 38, N°. 11, pages: 27-34, 2012.
- [24] A. Azeem, M. Sharif, M. Raza et M. Murtaza. A Survey: Face Recognition Techniques under Partial Occlusion. *International Arab Journal of Information Technology*, Vol. 11, No. 1, Janvier 2014.
- [25] W. W. Bledsoe. The model method in facial recognition. Panoramic Research Inc., Tech. Rep. PRI15, Palo Alto, CA, 1964.
- [26] T. Kenade. Picture Processing System by Computer Complex and Recognition of Human Faces. PhD thesis, Kyoto University, November 1973.
- [27] M. Kirby et L. Sirovich, Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, No. 1, pages: 103-108, January 1990.
-

-
- [28] M. Turk et A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pages: 71-86, 1991.
- [29] A. Ansari et M. Abdel-Mottaleb. 3D Face Modelling Using Two Orthogonal Views and a Generic Face Model. *Proceedings of International Conference and Multimedia and Expo*, Juillet 2003.
- [30] V. Blanz et T. Vetter. Face Recognition based on fitting 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, N°. 9, pages: 1063-1074, 2003.
- [31] X. Lu et A. K. Jain. Integrating Range and Texture Information for 3D Face Recognition. *Proceedings of IEEE WACV*, Breckenridge, Colorado 2005.
- [32] X. Lu et A. K. Jain. Matching 2.5D face scans to 3D models. *IEEE Transactions Pattern Analysis and Machine Intelligence*, Vol. 28, pages: 31- 43, Janvier 2006.
- [33] K. I. Chang, K. W. Bowyer et P. J. Flynn. An Evaluation of Multimodal 2D+3D Face Biometrics. *IEEE Transactions. on Pattern Analysis and Machine Intelligence*, Vol. 27, pages: 619-624, Avril 2005.
- [34] F. Tsalakanidou, S. Malassiotis et M. Srinatzis. Integration of 2D and 3D Images for enhanced face authentication. *Sixth International Conference on Automated Face and Gesture Recognition*, pages: 266-271, Mai 2004.
- [35] A. M. Bronstein, M. M. Bronstein et R. Kimmel. Three-dimensional face recognition. *International Journal of Computer Vision*, Vol.64, N°.1, pages: 5-30, Août 2005.
- [36] A. F. Abate, M. Nappi, D.Riccio et G. Sabatino. 2D and 3D face recognition: A survey. *Pattern Recognition Letters*, Vol. 28, pages: 1885–1906, 2007.
- [37] W. Zhao, R. Chellappa et A. Rosenfeld. Face recognition: a literature survey, *ACM Computing Surveys*, Vol. 35, pages: 399–458, Décembre 2003.
- [38] X. Tana, C. Songcan, Z. Zhou et F. Zhangb. Face recognition from a single image per person: A survey. *Pattern recognition*, Vol. 39, Issue 9, pages: 1725–1745, Septembre 2006.
- [39] R. Patel, N. Rathod et A. Shah. Comparative Analysis of Face Recognition Approaches: A Survey. *International Journal of Computer Applications*, Vol. 57, No.17, pages: 0975 – 8887, Novembre 2012.
- [40] S. Asht et R. Dass. Pattern Recognition Techniques: A Review. *International journal of Computer Science Telecommunications*, Vol. 3, Issue 8, pages: 25-29, Août 2012.
- [41] R. Dass, R. Rani et D. Kumar. Face Recognition Techniques: A Review. *International Journal of Engineering Research and Development*, Vol. 4, Issue 7, pages: 70-78, Novembre 2012.
- [42] V. Vijayakumari. Face Recognition Techniques: A Survey. *World Journal of Computer Application and Technology*, Vol. 1, N°. 2, pages: 41-50, 2013.
-

-
- [43] J. Yang et D. Zhang. Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, pages: 131- 137, 2004.
- [44] M. Visani, C. Garcia et C. Laurent. Comparing Robustness of Two-Dimensional PCA and Eigenfaces for Face Recognition. Dans *Proc. of the International Confer on Image Analysis and Recognition (ICIAR'04)*, Porto, Portugal, Octobre 2004.
- [45] J. Wu et Z. H. Zhou. Face recognition with one training image per person, *Pattern Recognition Letter*, Vol. 23, N°. 14, pages: 1711–1719, 2002.
- [46] B. Scholkopf, A. Smola, et K. Muller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, pages 1299–1319, 1998.
- [47] N. Sun, H. Wang, Z. Ji, C. Zou et L. Zhao. An efficient algorithm for Kernel two-dimensional principal component analysis. *Neural Computing & Applications*, Vol. 17, pages: 59-64, 2008.
- [48] Q. Yang et X. Q. Ding. Symmetrical Principal Component Analysis and Its Application in Face Recognition. *Chinese Journal of Computers*, Vol. 26, pages: 1146–1151, 2003.
- [49] A. P. Kumar, S. Das et V. Kamakoti. Face recognition using weighted modular principle component analysis. *Neural Information Processing*, Vol. 3316, Lecture Notes In Computer Science: Springer, pages: 362-367, Berlin / Heidelberg, 2004.
- [50] R. Gottumukkal et V. K. Asari, An improved face recognition technique based on modular PCA approach. *Pattern Recognition Letter*, Vol. 25, N°. 4, pages: 429–436, 2004.
- [51] K. R. Tan et S. C. Chen. Adaptively weighted subpattern PCA for face recognition. *Neurocomputing*, Vol. 64, pages: 505-511, 2005.
- [52] B. Moghaddam et A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, N°. 7, pages: 696–710, 1997.
- [53] P. N. Belhumeur, Joao P. Hespanha et D. Kreigman. Eigenfaces vs. Fisherfaces Recognition using class specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, N°. 7, pages: 711-720, 1997.
- [54] A. Martinez, A.C. Kak. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, N°. 2, pages: 228–233, 2001.
- [55] W. Zhao 2, R. Chellappa et P.J. Phillips. Subspace linear discriminant analysis for face recognition, Technical Report CAR-TR-914, Center for Automation Research, University of Maryland, 1999.
- [56] L.Chen, H.Liao, M.Ko, J.Lin et G.Yu. A New LDA-Based Face Recognition System which can Solve the Small Sample Size Problem. *Pattern Recognition*, 2000.
-

-
- [57] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, et K.R. Mullers. Fisher discriminant analysis with kernels. *Neural Networks for Signal Processing IX*, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop, pages: 41-48, Août 1999.
- [58] J. Lu, K.N. Plataniotis et A.N. Venetsanopoulos, Face recognition using kernel direct discriminant analysis algorithms, *IEEE Trans. Neural Networks* 14 (1), pages: 117–126, 2003.
- [59] M. Visani, C. Garcia et J.M. Jolion. Bilinear Discriminant Analysis for Face Recognition. Dans *Proceedings of the International Conference on Advances in Pattern Recognition (ICAPR 2005)*, Vol. 2, pages: 247-256, Août 2005.
- [60] H. Yu et J. Yang. A Direct LDA Algorithm for High-dimensional Data with Application to Face Recognition. *Pattern Recognition*, Vol.34, pages: 2067- 2070, 2001.
- [61] X. Wang et X. Tang. Dual-space Linear Discriminant Analysis for Face Recognition. Dans *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages: 564–569, 2004.
- [62] D. Zhou et X. Yang. Face Recognition Using Direct-Weighted LDA. Dans *8th Pacific Rim International Conference on Artificial Intelligence*, pages: 760-768, Auckland, New Zealand, 2004.
- [63] V. D. M. Nhat et S. Lee. Block LDA for Face Recognition. Dans *Computational Intelligence and Bioinspired Systems*. Vol. 3512, *Lecture Notes in Computer Science*: Springer, pages: 899-905, Berlin / Heidelberg, 2005.
- [64] C. Jutten et J. Herault. Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture. *Signal Process.*, Vol. 24, N°. 1, pages:1–10, 1991.
- [65] N. Kwak, C. H. Choi et N. Ahuja. Face recognition using feature extraction based on independent component analysis. Dans *Proceedings of IEEE International Conference on Image processing (ICIP 02)*, Vol. 2, pages: 337-340, Rochester, NY, USA, Septembre 2002.
- [66] M. Bartlett, J. Movellan et T. Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, Vol. 13, N°. 6, pages: 1450-1464, Novembre 2002.
- [67] B. A. Draper, K. Baek, M. S. Bartlett et J. R. Beveridge. Recognizing faces with PCA and ICA. *Comput. Vis. Image Underst.*, Vol. 91, pages: 115–137, Juillet 2003.
- [68] F. Bach et M. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, pages: 1-48, 2002.
- [69] M. Li, F. Wu et X. Liu. Face recognition based on WT, fastICA and RBF neural network. Dans *Third International Conference on Natural Computation*, Vol. 2, pages: 3-7, Août 2007.
-

-
- [70] Z. Mu-chun. Face recognition based on fastICA and RBF neural networks. Dans International Symposium on Information Science and Engineering, Vol. 1, pages: 588-592, Décembre 2008.
- [71] K.C. Kwak et W. Pedrycz. Face recognition using an enhanced independent component analysis approach. IEEE Transactions Neural Networks, 18(2), pages: 530-541, Mars 2007.
- [72] L. Zhang, Q. Gao et Z. D. Block independent component analysis for face recognition. Dans 14th International Conference on Image Analysis and Processing, pages: 217-222, Septembre 2007.
- [73] B. Luo, Y. J. Hao, W. H. Zhang et Z. S. Liu. Comparaison of PCA and ICA in face recognition. Dans International Conference on Apperceiving Computing and Intelligence Analysis, pages: 241-243, Chengdu, China, Décembre 2008.
- [74] N. Ahmed, T. Natarajan, et K. R. Rao. Discrete Cosine Transform. IEEE Transactions Computers, pages: 90-93, Janvier 1974.
- [75] Z. M. Hafed et M. D. Levine. Face recognition using discrete cosine transform. International Journal of Computer Vision, Vol. 43, N°. 3, pages:167–188, Juillet-Août 2001.
- [76] S. Wu, L. Wei, Z. Fang, R. Li et X. Ye. Infrared face recognition based on blood perfusion and sub-block dct in wavelet domain. Dans International conference on Wavelet Analysis and Pattern Recognition, 2007.
- [77] D. Omaia, V. D. Jankees Poel et V. Leonardo Batista. 2DDCT Distance based Face Recognition using a Reduced Number of Coefficients. Proceedings of Twenty Second Brazilian Symposium on Computer Graphics and Image Processing, pages: 291-298, 2009.
- [78] M. Azam et M.Y Javed. Discrete Cosine Transform (DCT) Based Face Recognition in Hexagonal Images. Proceedings of 2nd International Conference on Computer and Automation Engineering (ICCAE 2010), pages:. 168-170. Singapore, 26-28 Février 2010.
- [79] J. Shao, J. Jiang et X. Liu. Biomimetic Pattern Face Recognition Based on DCT and LDA. Artificial Intelligence and Computational Intelligence Lecture Notes in Computer Science, Vol. 7004, pages : 170-177, 2011.
- [80] R. Brunelli et T. Poggio. Face recognition through geometrical features. Proceedings of Second European Conference on Computer Vision (ECCV'92), pages: 792-800. Santa Margherita Ligure, Italy, 19–22 Mai, 1992.
- [81] J. Cox, J. Ghosn et P. N. Yianilos. Feature-Based Face Recognition Using Mixture-Distance. Dans Proceedings of Computer Vision and Pattern Recognition (CVPR '96), pages: 209 – 216. San Francisco, CA, USA, 18-20 Juin 1996.
-

-
- [82] Y. Ben Jemaa et S. Khanfir. Automatic local Gabor features extraction for face recognition. *International Journal of Computer Science and Information Security*, Vol. 3, No. 1, 2009.
- [83] B. S. Manjunath, R. Chellappa et C. von der Malsburg. A feature based approach to face recognition. Dans *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR' 1992)*, pages :373 – 378. Juin 1992.
- [84] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz et W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, Vol. 42, N°. 3, pages: 3003-11, Mars 1993.
- [85] L. Wiskott, J. M. Fellous, N. Kuiger et C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions PAMI*, Vol. 19, N° 7, pages: 775-779, Juin 1997.
- [86] L. Wiskott, J. M. Fellous, N. Kuiger et C. Malsburg. Face recognition by elastic bunch graph matching. *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, chapitre 11, pages: 355-396, 1999.
- [87] J. Zhang, Y. Yan et M. Lades. Face recognition: eigenface, elastic matching and neural nets. *Proceedings of the IEEE*, Vol. 85, Issue: 9, pages : 1423–1435, Septembre 1997.
- [88] C. Kotropoulos, A. Tefas, et I. Pitas. Morphological elastic graph matching applied to frontal face authentication under well-controlled and real conditions. *Pattern Recognition*, Vol. 33, pages: 1935-1947, 2000.
- [89] D. Monzo, A. Albiol et J. Sastre. HOG-EBGM vs. Gabor-EBGM. *15th IEEE International Conference on Image Processing (ICIP 2008)*, pages: 1636 – 1639, San Diego, CA, 12-15 Octobre 2008.
- [90] D. Monzo, A. Albiol et J.M. Mossi. Color HOG-EBGM for face recognition. *18th IEEE International Conference on Image Processing (ICIP 2011)*, pages: 785 – 788, Bruxel, 11-14 Septembre 2011.
- [91] L. Shen et L. Bai. A review on Gabor wavelets for face recognition. *Pattern Analysis and Applications*, Vol. 9, Issue 2-3, pages: 273–292, 18 Août 2006.
- [92] Y. Jin et Q. Ruan. Face Recognition Using Gabor-based improved supervised locality preserving projections. *Computing and Informatics*, Vol. 28, pages: 81–95, 2009.
- [93] T. Ahonen, A. Hadid, et M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, N°. 12, pages: 2037–2041, Décembre 2006.
- [94] S. Liao, X. Zhu, Z. Lei, L. Zhang et S. Z. Li. Learning mutli-scale block local binary patterns for face recognition. *Proceedings of International Conference on Biometrics (ICB 2007)*, pages: 828–837, Seoul, Korea, 27-29 Août, 2007.
-

-
- [95] D. Maturana, D. Mery et A. Soto. Face Recognition with Local Binary Patterns, Spatial Pyramid Histograms and Naive Bayes Nearest Neighbor classification. International Conference of the Chilean Computer Science Society (SCCC 2009), pages: 125 – 132, Santiago, TBD, Chile, 10-12 Novembre 2009.
- [96] N. S. Vu et A. Caplier. Face Recognition with Patterns of Oriented Edge Magnitudes. Computer Vision (ECCV 2010), tome 6311 de Lecture Notes in Computer Science, pages : 313–326. Springer Berlin / Heidelberg, 2010.
- [97] T. Do et E. Kijak. Face Recognition using co-occurrence histograms of oriented gradients. 37th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012), pages: 1301-1304, Japan, Mars 2012.
- [98] P. Penev et J. Atick. Local features analysis: A general statistical theory for object representation. *Computation in Neural Systems*, Vol. 7, No. 3, pages: 477–500, Août 1996.
- [99] V. Perlibakas. Face recognition using Principal Component Analysis and Log-Gabor Filters. Cornell University Library. Mai 2006.
- [100] S. Meshgini, A. Aghagolzadeh et H. Seyedarabi. Face Recognition Using Gabor Filter Bank, Kernel Principale Component Analysis and Support Vector Machine. *International Journal of Computer Theory and Engineering*, Vol. 4, N°. 5, Octobre 2012.
- [101] F. Bellakhddhar, K. Loukil et M. Abid. Face recognition approach using Gabor Wavelets, PCA and SVM. *International Journal of Computer Science Issues*, Vol. 10, Issue 2, N° 3, pages: 201-207, Mars 2013.
- [102] P. Viola et M. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR 2001)*. Vol. 1, pages: 511-518. 2001
- [103] P. Viola et M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, Vol. 57, N°. 2, pages: 137–154, 2004.
- [104] E. Hjelm et B. K. Low. Face Detection: A Survey. *Computer Vision and Image Understanding*, Vol. 83, pages: 236-274, 2001.
- [105] M-H. Yang. Recent Advances in Face Detection. *Proceedings of the 17th international Conference on Pattern Recognition*, 2004. ICPR 2004.
- [106] C. Zhang et Z. Zhang. A Survey of Recent Advances in Face Detection. Technical Report (2010), Microsoft Research, Microsoft Corporation One Microsoft Way.
- [107] G. Yang et T. S. Huang. “Human Face Detection in Complex Background”. *Pattern Recognition*, Vol. 27, Issue 1, pages: 53-63, Janvier 1994.
- [108] C. Kotropoulos et I. Pitas. Rule-based face detection in frontal views. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-97)*. Vol. 4, pages: 2537 – 2540. Munich, 21-24 Avril 1997
-

-
- [109] A. Schwaninger, S. Ryfb et F. Hoferb. Configural information is processed differently in perception and recognition of faces. *Vision Research*, Vol. 43, Issue 14, pages 1501–1505, Juin 2003.
- [110] P. Sinha. Object Recognition via Image Invariants: A Case Study. *Investigative Ophthalmology and Visual Science*, Vol. 35, No. 4, pages: 1735-1740, 1994.
- [111] J. Wang et H. Yang. Face Detection Based on Template Matching and 2DPCA Algorithm. *Congress on Image and Signal Processing*, pages: 575-579, Sanya, Chine, Mai 2008.
- [112] H. Pan, Y. Zhu et L. Xia, Efficient and accurate face detection using heterogeneous feature descriptors and feature selection, *Computer Vision and Image Understanding*, Vol. 117, pages: 12–28, 2013.
- [113] N. Shilbayeh et G. Al-Qudah. Face Detection System Based On MLP Neural Network. *Recent Advances in Neural Networks, Fuzzy Systems & Evolutionary Computing*, pages: 238-243, 2008, ISBN: 978-960-474-195-3.
- [114] K. Prakash. Face Detection using Neural Network. *International Journal of Computer Applications*, Vol. 1, No. 14, pages: 36-39, 2010.
- [115] A. Kaushal et J. P. Raina. Face Detection using Neural Network & Gabor Wavelet Transform. *International Journal of Computer Science and Technology*, Vol. 1, Issue. 1, pages: 58-63, Septembre 2010.
- [116] M. Abadi. Face Detection with the Help of Gabor Wavelets Characteristics and Neural Network Classifier. *American Journal of Scientific Research*, Issue. 36, pages: 67-76, 2011.
- [117] O. N. A. AL-Allaf. Review of face detection systems based artificial neural networks algorithms. *The International Journal of Multimedia & Its applications (IJMA)*, Vol. 6, No.1, Février 2014.
- [118] E. Osuna, R. Freund et F. Girosi. Training Support Vector Machines: An Application to Face Detection. Dans *Proceedings of IEEE Conference Computer Vision and Pattern Recognition (CVPR'97)*, pages: 130-136, 17-19 Juin 1997, Puerto Rico.
- [119] H-J. Lin, S-H. Yen, J-P. Yeh et M-J. Lin. Face Detection Based on Skin Color Segmentation and SVM Classification. Dans *Second International Conference on Secure System Integration and Reliability Improvement (SSIRI '08)*, pages: 230 – 231, 14-17 Juillet 2008, Yokohama.
- [120] S. Ravi et S. Wilson. Face detection with facial features and gender classification based on support vector machine. *International Journal of Imaging Science and Engineering*, 2010, Special Issue.
- [121] J. Mingxing, D. nqiang, C. Tao et Y. Ning. An improved detection algorithm of face with combining adaboost and SVM. *25th Control and Decision Conference (CCDC 2013)*, pages: 2459 – 2463, Chine, 25-27 Mai 2013.
-

-
- [122] W. Zheng et S. M. Bhandarkar. Face detection and tracking using a Boosted Adaptive ParticleFilter. *Journal of Visual Communication and Image Representation*. Vol. 20, Issue 1, pages 9–27, Janvier 2009.
- [123] B. Wu, H. Ai, C. Huang et S. Lao. Fast rotation invariant multi-view face detection based on real adaboost. *IEEE International Conference on Automatic Face and Gesture Recognition*, pages: 79–84, 2004.
- [124] J. Zeng, X. Cao et J. Gan. An Improvement of AdaBoost for Face Detection with Random Forests. *Advanced Intelligent Computing Theories and Applications Communications in Computer and Information Science Volume 93*, pages: 22-29, 2010.
- [125] L. Xue et Z. Li, Using Skin Color and HAD-AdaBoost Algorithm for Face Detection in Color Images. *National Conference on Information Technology and Computer Science (CITCS 2012)*, pages: 233-236. Published by Atlantis Press.
- [126] T. Huang et Z. Wang. Face Detection By Improved AdaBoost. *2nd International Conference on Computer Science and Network Technology*. Pages: 981- 985. CHANGCHUN, CHINA, 2012.
- [127] P. S. Hiremath et P. Danti. A Detection of multiple faces in an image using skin color information and lines-of-separability face model. *International Journal of Pattern recognition and Artificial Intelligence*. Vol. 20, Issue 01, Février 2006.
- [128] A. Soetedjo et K. Yamada. Skin Color Segmentation Using Coarse-to-Fine Region on Normalized RGB Chromaticity Diagram for Face Detection. *IEICE TRANSACTIONS on Information and Systems*, Vol. E91-D, No.10, pages: 2493-2502, Octobre 2008.
- [129] K. Hawari, J. Ma et R. Xiao. An Innovative Face Detection based on Skin Color Segmentation, *International Journal of Computer Applications (0975 – 8887)*, Vol. 34, No.2, pages: 6-10, Novembre 2012.
- [130] Y-H. Chen, K-T Hu et S-J. Ruan, Statistical skin color detection method without color transformation for real-time surveillance systems. *Engineering Applications of Artificial Intelligence*, Vol. 25, pages: 1331–1337, 2012.
- [131] M. H. Rahman et J.Afrin. Human Face Detection in Color Images with Complex Background using Triangular Approach. *Global Journal of Computer Science and Technology Graphics & Vision*, Vol. 13, Issue 4, pages: 45-50, 2013.
- [132] D. Ghimire et J. Lee. A Robust Face Detection Method Based on Skin Color and Edges. *Journal of Information Processing Systems*, Vol.9, No.1, pages: 141-156, Mars 2013.
- [133] M. F. Augusteijn et T. L. Skujca. Identification of Human Faces through Texture-Based Feature Recognition and Neural Network Technologie. Dans *Proceedings of the IEEE Conference on Neural Networks*, pages: 392-398, 1993.
-

-
- [134] R. Herpers. Detection and tracking of faces in real environments. In Proceedings of the IEEE International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems, pages 96–104, 1999.
- [135] J. Wang et T. Tan. A new face detection method based on shape information. *Pattern Recognition Letters*, vol. 21, pages: 463–471, 2000.
- [136] V. Vezhnevets, V. Sazonov et A. Andreeva. A Survey on Pixel-Based Skin Color Detection Technique, Dans Proceedings of the GraphiCon 2003, pages: 85-92. 2003
- [137] J. Brand et J.S. Mason. A comparative assessment of three approaches to pixel-level human skin-detection. Dans Proceedings 15th International Conference on Pattern Recognition, Vol. 1, pages: 1056-1059, Barcelona, Spain, September 2000.
- [138] P. Peer, J. Kovac et F. Solina. Human skin colour clustering for face detection. International Conference on Computer as a Tool, EUROCON 2003, Ljubljana, Slovenia, Septembre 2003.
- [139] M. Harville, H. Baker, N. Bhatti et S. Susstrunk. Consistent Image-Based Measurement and Classification of Skin Color. International Conference on Image Processing, pages 11–14, 2005.
- [140] S.L. Phung, A. Bouzerdoum et D. Chai. Skin segmentation using color pixel classification: Analysis and comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, N^o.1, pages: 148–154, 2005.
- [141] N. A. Hikal et R. Kountchev. Skin color segmentation using adaptive PCA and modified elliptic boundary model. International Conference on Advanced Computer Science and Information System (ICACSIS 2011), pages: 407 – 412, Jakarta, 17-18 Décembre 2011.
- [142] S. Derivaux, S. Lefèvre, C. Wemmert et J.J. Korczak. Segmentation par ligne de partage des eaux basées sur des connaissances texturales. In Colloque GRETSI sur le Traitement du Signal et des Images, Troyes, France, Septembre 2007.
- [143] V. Grau, A.U.J. Mewes, M. Alcaniz, R. Kikinis et S.K. Warfield. Improved watershed transform for medical image segmentation using prior information. *IEEE Transactions on Medical Imaging*, Vol. 23, N^o. 4, pages: 447–458, Avril 2004.
- [144] S. Beucher et C. Lantuéjoul. Use of watersheds in contour detection. International workshop on image processing, real-time edge and motion detection. Rennes, France, Septembre 1979.
- [145] F. Meyer et S. Beucher. Morphological segmentation. *Journal of visual communication and image representation*. Vol. 1, N^o. 1, pages: 21-46, 1990.
- [146] L. Vincent et P. Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13(6), pages : 583–598, Juin 1991.
-

-
- [147] P. F. Felzenszwalb et D. P. Huttenlocher. Distance Transforms of Sampled Functions. *THEORY OF COMPUTING*, Vol. 8, pages: 415–428, 2012.
- [148] J. G. Daugman. Complete discrete 2D Gabor transforms by neural network for image analysis and compression. *IEEE Transactions on acoustics, Speech, and Signal Processing*. Vol. 36, N°. 7, pages: 1169-1179, Juillet 1988.
- [149] E. M. Chutorian et M. M. Trivedi. Head Pose Estimation in Computer Vision : A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, pages: 607– 626, 2009.
- [150] M. Feuilloley. Étude d’algorithmes d’apprentissage artificiel pour la prédiction de la syncope chez l’homme. Thèse de doctorat, Université d’Angers, 2009.
- [151] J. P. Rennard. Réseaux neuronaux : une introduction accompagnée d'un modèle Java. Vuilbert, Paris, 2006.
- [152] M. Rizon, M. F. Hashim, P. Saad et S. Yaacob. Face Recognition using Eigenfaces and Neural Networks. *American Journal of Applied Sciences*, Vol. 3, N° 6, pages: 1872-1875, 2006.
- [153] M. S. S. Ranawade. Face Recognition and Verification Using Artificial Neural Network. *International Journal of Computer Applications*, Vol. 1, N°. 14, pages: 23-30, 2010.
- [154] M. Agarwal. Face Recognition Using EigenFaces and Artificial Neural Network. *International Journal of Computer Theory and Engineering*, Vol. 2, No. 4, pages: 624-629, Aout 2010.
- [155] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag New York. 1995.
- [156] S. Kim, Y. J. Park, K. Toh et S. Lee. SVM-based feature extraction for face recognition. *Pattern Recognition*, Vol. 43, Issue 8, pages 2871–2881, Août 2010.
- [157] J. Hongjun Jia et A. M. Martinez. Support Vector Machines in face recognition with occlusions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009)*. Pages: 136 – 141, Miami, FL, 20-25 Juin 2009.
- [158] M.O. Faruqe. Face recognition using PCA and SVM. Dans *3rd International Conference on Anti-counterfeiting, Security, and Identification in Communication, 2009 (ASID 2009)*, pages: 97 – 101, Hong Kong, 20-22 Août 2009.