

078



Mémoire de magister

Présenté à l'Université de Guelma
Faculté des Sciences et de l'Ingénierie

Département de Génie Electrique
Spécialité : Informatique Industrielle et Imagerie

Présenté par M^{lle} **Rouabhia Chahrazed**

Approches Bidimensionnelles Hybride Neuro-ACPD2D & Locales pour la Reconnaissance Automatique de Visages

Sous la direction du **Pr. Hicham Tebbikh**

Jury

Pr. M. NEMAMCHA
Pr. H. TEBBIKH
Pr. N. DOGHMANE
Dr. M. ROUAINIA
Dr. H. SERIDI

Université de Guelma
Université de Guelma
Université de Annaba
Université de Skikda
Université de Guelma

Président
Rapporteur
Examineur
Examineur
Examineur

11 Mars 2007



C'est absurde, repartit Ulrich avec force.

J'ai dit que ce qui comptait, ce n'était pas un faux pas, mais le pas qui suit ce faux pas. Mais qu'est-ce qui compte après le pas suivant ? Sans doute, bien sûr, celui qui suit ? Et après le énième pas, le pas n+1 ?

Cet homme devrait donc vivre privé de fin et de décision, privé même, somme toute, de réalité. Pourtant il est bien vrai que c'est toujours le pas suivant qui compte.

La vérité est que nous ne disposons d'aucune méthode pour traiter comme il faudrait cette série infatigable.

Robert Musil,
L'homme sans qualités

Remerciements

*Louange à ALLAH Qui m'a aidé à atteindre ce but et défier tous les obstacles
afin de compléter ce mémoire de magister.*

Je désire exprimer vivement ma reconnaissance à l'égard de mon directeur de mémoire, le professeur **Tebbikh Hicham**, directeur du Laboratoire d'Automatique et Informatique de Guelma - LAIG - pour l'opportunité qu'il m'a offert de travailler sous sa supervision, son aide, ses précieux conseils et sa confiance dont j'ai bénéficié durant la réalisation de ce travail de recherche.

J'adresse mes sincères remerciements aux membres du jury :

Monsieur **Nemamcha Mohamed**, professeur et recteur de l'université de Guelma, pour m'avoir fait l'honneur de présider mon jury de mémoire,
Messieurs **Doghmane Noureddine**, professeur à l'université de Annaba et **Séridi Hamid**, docteur à l'université de Guelma et Madame **Rouaïnia Mounira**, docteur à l'université de Skikda, qui ont accepté de lire et de juger mon travail malgré leurs responsabilités et emplois du temps chargés.

Je suis également très reconnaissante envers messieurs **Aissaoui** et **Séridi** pour l'aide qu'ils m'ont apportée. Qu'ils trouvent ici l'expression de ma gratitude.

Merci à ma famille qui m'a supporté patiemment, dans tous les sens du terme, tout au long du chemin.

Rouabhia Chahrazed

Résumé

En plus des variations d'illumination, de la pose de la tête, des expressions faciales, des occultations partielles, et du vieillissement,... les approches globales, utilisant le visage entier pour la reconnaissance, souffrent de la grande taille de données à traiter due à la transformation des images de visages en vecteurs. Pour corriger ce problème majeur, nous proposons une approche hybride bidimensionnelle **ACPDL2D**, intégrée dans un système automatique de reconnaissance faciale comportant trois étapes :**i)** localisation automatique de visages, **ii)** extraction de caractéristiques et réduction de dimensions par l'**ACPDL2D** et **iii)** une classification par un **PMC**. Le système proposé **Neuro-ACPDL2D** assure ainsi un codage et une compression de données très efficace, via une projection hybride bilatérale dans un espace bidimensionnel, dépassant celle des autres méthodes récemment proposées tout en garantissant une reconnaissance satisfaisante.

Dans le but de corriger le problème des occultations partielles, nous avons également proposé des **approches locales bidimensionnelles** avec une classification neuronale. Elles reposent sur la forme des éléments faciaux au sens où elles s'appliquent, séparément, sur deux régions faciales (une contenant les yeux avec les sourcils et l'autre le nez), après une localisation automatique de ces derniers. Ces approches sont plus performantes que les approches géométrique et globales bidimensionnelles en termes d'espace mémoire et de reconnaissance, et elles sont privilégiées car invariantes aux arrières plans, au style de coiffure, moins sensibles aux variations d'illumination et présentent l'avantage de pouvoir identifier une personne dont le visage est en partie cachée.

Mots clés : ACP2D, ADL2D, Approches locales bidimensionnelles, Localisation, Éléments faciaux, Neuro-ACPDL2D, PMC, Reconnaissance automatique de visages, Séquences vidéo.

Abstract

In addition to the variations of illumination, pose of the head, facial expressions, partial occlusion, and age,... the global approaches, using the whole face for recognition, suffer from the big size of data due to the transformation of the faces images into vectors. To correct this major problem, we propose a hybrid two-dimensional approach called **ACPDL2D**, integrated in an automatic face recognition system that contains three stages: i) automatic localization of faces, ii) features extraction and dimension reduction by the ACPDL2D and a iii) classification by a MLP. The system proposed ensures a coding and a very effective data compression, via a bilateral hybrid projection in a two-dimensional space, exceeding the other methods that are recently proposed and it guarantees a satisfactory recognition rate.

To correct the problem of partial occlusions, we have also, proposed a two-dimensional local approaches with neural classification. They are based on the shape of the facial features and applied on two local regions (one contains eyes with the eyebrows and the other contains nose). These approaches are more powerful than the geometric and the two-dimensional global approaches in terms of memory capacity and recognition, and they are privileged because they are invariant to the background, hair style, less sensitive to the variations of illumination and have the advantage of being able to identify a person's face which is partially occult.

Key words: Automatic face recognition, Facial features, Localization, MLP, Neuro-ACPDL2D, 2DLDA, 2DPCA, Two-dimensional local approaches, Video sequences.

الملخص

بالإضافة إلى تغييرات الإضاءة، وضعية الرأس، تعبيرات الوجه، التغطيات الجزئية وتقدم السن،... فإن الطرق الكلية لمعرفة الوجوه تعاني من مشكلة الحجم الكبير للمعطيات نتيجة تحويل المصفوفات إلى أشعة. لاستخراج المميزات وتقليص الأبعاد نقترح، في إطار هذه المذكرة، طريقة مختلطة ثنائية الأبعاد مدمجة في نظام آلي لمعرفة الوجوه يتضمن المراحل التالية: أ) التعهيد الآلي للوجوه ، ب) استخراج المميزات وتقليص الأبعاد بطريقة مختلطة تجمع بين التحليل بالمركبات الأساسية والتحليل المميز الخطي و ج) تصنيف الوجوه بالشبكة العصبونية متعددة الطبقات. مقارنة مع طرق أخرى حديثة فإن الطريقة المقترحة توفر تقليصا ممتازا لحجم المعطيات مع المحافظة على نسبة معرفة حسنة.

لمعالجة مشكلة التغطيات الجزئية قمنا باقتراح الطرق الموضعية ثنائية الأبعاد نسبة لأنه بعد التعهيد الآلي للحاجبين، العينين و الأنف نطبق الطرق الكلية ثنائية الأبعاد على المناطق الموضعية المتضمنة لهذه العناصر. مقارنة مع الطريقة الهندسية والطرق الكلية، فإن الطرق الموضعية ثنائية الأبعاد قدمت أحسن النتائج إضافة إلى كونها أقل تأثرا بتغيرات الإضاءة والخلفيات، تعالج معطيات أقل حجما من الطرق الكلية وتمكن من التعرف على الوجوه المغطاة جزئيا.

الكلمات الدالة: الطرق الكلية، الطرق الموضعية ثنائية الأبعاد، عناصر الوجه، المعرفة الآلية للوجوه، صور فيديو متتابعة.

Sommaire

Introduction Générale	1
Chapitre I	
Reconnaissance de Formes	
Introduction	3
I Reconnaissance de formes (RdF)	4
I.1 Histoire et définition	4
I.2 Approches de la reconnaissance de formes	4
I.2.1 Appariement de gabarit	4
I.2.2 Approche statistique	4
I.2.3 Approche syntaxique	5
I.2.4 Réseaux de neurones	5
I.3 Système de reconnaissance de formes	5
I.3.1 Acquisition et représentation de données	6
I.3.2 Génération de caractéristiques	6
I.3.3 Extraction de caractéristiques et réduction de dimensions	6
I.4 Différentes approches de la reconnaissance de formes statistique	6
II Classification	7
II.1 Représentation de données	7
II.2 Classification et classificateur	8
II.2.1 Classification non supervisée	9
II.2.2 Classification supervisée	10
II.3 Performances d'un classificateur	10
II.4 Combinaison de classificateurs	11
II.5 Théorie de la décision	11
II.5.1 Règle de décision optimale de Bayes	11
II.5.2 Erreur de classification (probabilité d'erreur)	12
II.6 Le plus proche voisin	13
II.7 Applications de la règle du plus proche voisin pour la reconnaissance de visages	14
III Réduction de dimensions	15

III.1 Méthodes d'extraction de caractéristiques	15
III.2 Analyse en composantes principales (ACP)	16
III.2.1 Principe et algorithme	17
III.2.2 Propriétés	18
III.2.3 Inconvénients	18
III.3 Analyse discriminante linéaire (ADL)	19
III.3.1 Principe	19
III.3.2 Calcul de la matrice optimale de projection	20
III.3.3 Algorithme	20
III.4 Méthodes d'extraction de caractéristiques et de réduction de dimensions appliquées dans la reconnaissance de visages	22
Conclusion	22
Chapitre II	
Réseaux de Neurones Artificiels	
Introduction	23
I Généralités sur les Réseaux de Neurones Artificiels	24
I.1 Neurone biologique	24
I.1.1 Structure	24
I.1.2 Fonctionnement	25
I.2 Neurone formel	25
Interprétation géométrique du fonctionnement du neurone formel	26
I.3 Perceptron	27
Limitation du perceptron	28
I.4 ADALINE et MADALINE	29
I.5 Réseaux de neurones artificiels	31
I.5.1 Définition	31
I.5.2 Architecture	31
I.5.2.1 Réseaux de neurones bouclés	31
I.5.2.2 Réseaux de neurones non bouclés	31
I.5.2.3 Réseaux de neurones multicouche	32
I.5.2.4 Séparation linéaire par un réseau de neurones multicouche	33
I.6 Apprentissage des réseaux de neurones artificiels	33
I.6.1 Apprentissage supervisé	34
I.6.2 Apprentissage renforcé	34
I.6.3 Apprentissage non supervisé	34
I.7 Fonctions d'activation	34
II Perceptron Multi-Couche (PMC)	35
II.1 Présentation du réseau	35
II.2 Apprentissage du PMC	36
Algorithme de rétropropagation de l'erreur	37
II.3 Problèmes de l'algorithme de rétropropagation de l'erreur	38

II.3.1 Minima locaux	38
II.3.2 Lenteur de la convergence	38
II.3.3 Critère d'arrêt	39
II.4 Saturation des neurones	39
II.5 Généralisation	40
II.6 Choix des paramètres du PMC	40
II.6.1 Fonction d'activation	40
II.6.2 Initialisation des poids	41
II.6.3 Taux d'apprentissage	43
II.6.4 Taille du perceptron multicouche	43
II.7 Applications des réseaux de neurones pour la détection et la reconnaissance de visages	44
	45
Conclusion	
Chapitre III	
Biométrie & Reconnaissance de Visages	
Introduction	46
I Biométrie & Identification d'Individus	47
I.1 Identification d'individus	47
I.2 Biométrie	47
I.2.1 Définition	47
I.2.2 Domaines d'application	47
I.3 Systèmes biométriques	48
I.4 Classification des systèmes de reconnaissance biométrique	49
I.5 Exemples de caractéristiques biométriques	50
I.5.1 L'iris	51
I.5.2 La rétine	51
I.5.3 La géométrie de la main	51
I.5.4 Les empreintes digitales	51
I.5.5 La voix	52
I.5.6 La signature	52
I.5.7 Le thermogramme facial	52
I.5.8 Le visage	52
I.6 Multimodalité	53
I.6.1 Différentes formes de multimodalité	53
I.6.2 Fusion de l'information dans les systèmes biométriques multimodaux	53
I.7 Bases de données multimodales	54
II Reconnaissance de Visages (RdV)	54
II.1 Historique de la reconnaissance de visages	54
II.2 Visage humain comme stimuli	55
II.3 Psychologie / Neurosciences et reconnaissance de visages	55
II.4 Système de reconnaissance automatique de visages	56
II.4.1 Définition	56
II.4.2 Reconnaissance de visages à partir de séquences d'images vidéo	57
II.4.3 Détection et localisation de visages	57

II.4.4	Détection des éléments caractéristiques du visage	58
II.4.5	Suivi de visages et d'éléments faciaux	58
II.4.6	Reconnaissance d'expressions faciales	58
II.5	Approches de la reconnaissance de visages	59
II.5.1	Approches basées sur les éléments caractéristiques du visage	59
II.5.2	Approches globales	60
II.5.2.1	Eigenfaces	60
a.	Algorithme	60
b.	Limitations	61
c.	Choix du nombre de vecteurs propres à retenir	62
II.5.2.2	Analyse discriminante linéaire (ADL)	63
II.5.2.3	Fisherfaces / Algorithme	64
II.5.3	Approches hybrides	64
II.6	Problèmes d'illumination et de pose	65
II.7	Extraction des éléments caractéristiques du visage	67
II.8	Evaluation des systèmes de reconnaissance automatique de visages	69
Conclusion		70
Chapitre IV		
Approches Proposées & Résultats		
Introduction		71
I Approche Hybride Bidimensionnelle Neuro-ACPDL2D et Approches Locales Bidimensionnelles		72
I.1 Méthodes bidimensionnelles d'extraction de caractéristiques		72
I.1.1 Analyse en composantes principales bidimensionnelle (ACP2D)		72
I.1.1.1	Principe	73
I.1.1.2	Reconstruction d'images originales	74
I.1.1.3	Classification de visages	75
I.1.2 Analyse discriminante linéaire bidimensionnelle (ADL2D)		75
I.1.2.1	Analyse discriminante linéaire bidimensionnelle orientée en lignes (ADL2DoL)	75
I.1.2.2	Analyse discriminante linéaire bidimensionnelle orientée en colonnes (ADL2DoC)	76
I.2 Supériorité de l'ACP2D et de l'ADL2D		76
I.3 Inconvénients de l'ACP2D et de l'ADL2D		76
I.4 L'approche Hybride Bidimensionnelle proposée : Neuro-ACPDL2D		78
I.4.1	Idée et motivations	78
I.4.2	Localisation automatique de visages	78
I.4.3	Extraction de caractéristiques et réduction de dimension par l'ACPDL2D	79
I.4.4	Classification de visages	80
I.4.5	Organigramme et algorithme de l'approche Neuro-ACPDL2D	81
I.5 Approches locales bidimensionnelles proposées : Yeux-ACP2D, Yeux-ADL2DoC et Yeux-ACPDL2D & Nez-ACP2D, Nez-ADL2DoC et Nez-ACPDL2D		83
I.5.1	Idée et motivations	83
I.5.2	Système proposé	84

I.5.3 Localisation automatique des éléments caractéristiques du visage	85
I.5.3.1 Les yeux	85
I.5.3.2 Les sourcils	85
I.5.3.3 Le nez et la bouche	85
I.5.4 Extraction des régions contenant les éléments du visage	87
II Résultats Expérimentaux et Interprétations	89
II.1 Résultats expérimentaux des méthodes globales	91
II.1.1 Présentation de la base de visages utilisée	92
II.1.2 Résultats de la localisation automatique de visages	92
II.1.3 Effet du choix de la topologie du PMC	92
II.1.4 Reconnaissance de visages par les méthodes globales bidimensionnelles	93
II.1.4.1 Résultats de Neuro-ACP2D	93
II.1.4.2 Résultats de Neuro-ADL2DoC	94
II.1.4.3 Résultats de l'approche hybride Neuro-ACPD2D	94
II.1.5 Comparaison des résultats	95
II.2 Reconnaissance de visages par les méthodes locales bidimensionnelles	95
II.2.1 Applications sur les régions contenant les sourcils et les yeux	95
II.2.1.1 Résultats de Yeux-ACP2D	95
II.2.1.2 Résultats de Yeux-ADL2DoC	96
II.2.1.3 Résultats de Yeux-ACPD2D	96
II.2.1.4 Comparaison des résultats	96
II.2.2 Applications sur les régions contenant les nez	97
II.2.2.1 Résultats de Nez-ACP2D	97
II.2.2.2 Résultats de Nez-ADL2DoC	98
II.2.2.3 Résultats de Nez-ACPD2D	98
II.2.2.4 Comparaison des résultats	98
II.2.3 Comparaison entre les résultats obtenus par les méthodes globales et locales bidimensionnelles	99
Conclusion	99
Conclusion Générale	101
Annexe A	102
Annexe B	103
Annexe C	104
Annexe D	106
Bibliographie	108

Introduction Générale

Etes-vous réellement X ? Pour le prouver, il suffit de présenter une pièce d'identité telle que le passeport. Cette pièce comporte en plus de votre nom et votre prénom, vos caractéristiques physiques: le visage, la taille, la couleur des cheveux et des yeux et une empreinte digitale, la caractéristique la plus importante car la plus fiable.

Mais avant la disponibilité de telles pièces d'identité, comment les individus étaient identifiés ? Mesurer et comparer, était la première méthode scientifique, si on peut dire, d'identification de personnes. Elle était créée et appliquée en 1880 par Alphonse Bertillon, créateur de la police scientifique, à des fins d'identification des délinquants. En utilisant 11 mesures anatomiques (longueur et largeur de la tête, longueurs des pieds, longueur du majeur,...), cet employé de la préfecture de police de Paris, parvenait à connaître des récidivistes en photographiant les personnes et notant leurs mesures prises sur leurs photos et sur des fiches. Cette méthode a été appliquée dans la police jusqu'à la découverte, par Francis Galton, de la permanence et de l'unicité des empreintes digitales. Cette découverte a mis fin au bertillonnage mais la naissance de l'identification à partir des empreintes digitales, adoptées par Scotland Yard dès 1901 et la préfecture de Paris dès 1903. Cependant, plusieurs auteurs s'accordent à faire remonter l'utilisation des empreintes digitales par les chinois et les japonais, à un millénaire.

Il n'y a pas que les empreintes digitales qui sont propres à chaque individu, le visage, la voix, la démarche, la forme de la main, l'ADN,... sont également uniques pour chacun de nous. La technologie appliquant ces caractéristiques, dites biométriques, à des fins de reconnaissance est la «biométrie», dont le seul but est de faciliter la vie et sécuriser les personnes et les endroits. Parmi toutes ces caractéristiques, nous intéressons au visage dans le cadre d'une reconnaissance automatique faciale. A cet objectif, nous lui donnons une attention particulière étant l'élément de base de ces systèmes. Un objet visuel pas comme les autres, le visage caractérise, tout seul, une personne : son identité, son état affectif, ... c'est pourquoi, on le trouve dans les anciennes gravures, dans les pièces de monnaies, dans les timbres postaux,... d'ailleurs le premier timbre émis dans le monde, en 1840, portait le visage de la reine de Grande-Bretagne, Victoria.

Le visage a fasciné, bien avant, les scientifiques, des peintres, des sculpteurs et des psychologues,... Ces derniers suggèrent que les nourrissons naissent avec l'habilité de reconnaître les visages ; ils préfèrent suivre les visages que d'autres objets et distinguent le visage de leur mère des autres. Cette habilité cognitive se développe en grandissant et l'adulte devient capable de reconnaître, non seulement, les personnes par leur visages, mais aussi, leur état émotionnel.

De leur côté, les peintres et les sculpteurs, dans leurs recherches, pour décrire la forme idéale du corps humain, établissaient des règles servant à déterminer les proportions idéales

du corps humain y compris le visage : les canons de beauté. Les travaux de Léonard de Vinci et de Michelangelo se sont inspirés de ces canons et leurs chef-d'œuvres en témoignent.

Quant aux ingénieurs, inspirés du système de perception humain, ils tentaient de créer des machines dotées de cette capacité. Depuis le premier système automatique de T.Kanade datant de 1973 jusqu'aujourd'hui, le domaine de la reconnaissance automatique faciale s'enrichit de jour en jour par de nouvelles méthodes et /ou de méthodes améliorées dans le seul but d'approcher le système visuel humain.

De notre côté, nous avons essayé de contribuer dans ce domaine en proposant d'une part une approche pour corriger un problème majeur, rencontré en appliquant les méthodes globales; celui de la grande taille des images de visage. Cette approche hybride **Neuro-ACPDL2D** effectue une compression de données très efficace via une projection hybride bilatérale bidimensionnelle. D'autre part, nous proposons, pour corriger le problème des occultations partielles, des **approches locales bidimensionnelles** qui reposent uniquement sur la forme des éléments faciaux. Ces approches sont invariantes aux arrières plans, moins sensibles aux variations d'illumination et présentent l'avantage de pouvoir identifier une personne dont le visage est en partie cachée. Les méthodes proposées sont intégrées dans un système de reconnaissance automatique de visages comportant, pour Neuro-ACPDL2D, une étape de localisation automatique de visages et pour les méthodes locales bidimensionnelles, une localisation automatique des éléments faciaux et la classification, dans les deux cas, est assurée par un perceptron multicouche (PMC) entraîné par l'algorithme «Backpropagation».

Le fruit de notre recherche bibliographique ainsi que la méthodologie des approches proposées sont répartis en quatre chapitres composant ce mémoire ; après l'introduction générale, le **1^{er} chapitre** présente des généralités sur la reconnaissance de formes et la classification avec un espace réservé aux méthodes d'extraction de caractéristiques et de réduction de dimensions appliquées en reconnaissance de visages.

Le **2^{ème} chapitre** traite les réseaux de neurones artificiels, dans les paragraphes de sa première partie, nous rappelons leurs principes de base notamment pour la classification et dans la deuxième partie, nous décrivons à partir d'une bibliographie détaillée, un classificateur très utilisé ; le perceptron multicouche (PMC) en se basant sur les méthodes qui déterminent sa topologie et ses paramètres optimaux.

Quant au **3^{ème} chapitre**, il est dédié à l'identification de personnes ; il donne un aperçu de quelques techniques biométriques et des systèmes multimodaux avant de détailler un système biométrique très prometteur car basé sur une caractéristique naturelle utilisée quotidiennement pour reconnaître les personnes; c'est le système de la reconnaissance faciale. Une analyse et revue de littérature sont exposées permettant de présenter les différentes approches proposées avec leurs avantages et limites et de démontrer l'avancement des recherches dans le domaine de la reconnaissance automatique de visages.

Le **4^{ème} chapitre** est le cœur de ce mémoire car il détaille le principe des approches proposées, les différentes étapes de leurs mises au point et les résultats de simulations et leurs interprétations. Les tests ont été menés sur la base de visages B-LAIG acquise au LAIG en 2003/2004, ils montrent clairement la supériorité des approches locales bidimensionnelles vis-à-vis les approches globales bidimensionnelles et la méthode géométrique Neuro-ACP proposée dans [69].

Enfin, nous concluons ce mémoire par des remarques et perspectives des futures recherches.

Chapitre I

Reconnaissance de Formes

Introduction

En s'inspirant des phénomènes naturels, les chercheurs ont toujours tenté, et tenteront, d'imiter les capacités des humains à reconnaître les objets. Ils arrivaient à réaliser des machines, qualifiées d'intelligentes, qui peuvent apprendre et s'adapter face aux différentes situations. Ainsi la reconnaissance de formes permet à ces machines d'effectuer une perception aussi proche que celle de l'être humain : détecter un défaut dans un tissu, une cellule malade, reconnaître une signature, un visage dans une foule, suivre un missile,...

Ce chapitre débute par rappeler, brièvement, les notions de bases de la reconnaissance des formes et de la classification avant de réserver un espace aux méthodes d'extraction de caractéristiques et de réduction de dimensions appliquées dans la reconnaissance de visages, notre axe de recherche. En particulier, l'analyse en composantes principales et l'analyse discriminante linéaire seront détaillées.

I. Reconnaissance de Formes (RdF)

I.1 Histoire et définitions:

Avant de parler de la reconnaissance de formes (RdF), il est important de définir la forme (The pattern), pour ce faire nous allons adopter, après sa traduction, la définition donnée par Watanabe en 1985 et citée dans [1]: "la forme est le contraire du chaos, c'est une entité largement définie que l'on peut nommer". La forme peut être donc un signal de parole, une image médicale, un visage humain, des empreintes digitales, un document texte,... et la reconnaissance de formes a pour but d'identifier ces formes en se basant sur certaines de leurs caractéristiques pour enfin les regrouper dans des catégories appelées classes.

Historiquement, la reconnaissance de formes remonte aux années 40 [3] avec des travaux de détection effectués dans le traitement du signal et qui ont servi de base pour le développement de nombreuses méthodes utilisant d'autres données que les signaux. Jusqu'aux années 60 et 70, la RdF statistique [1] se reposait sur des techniques simples qui étaient appliquées à un nombre limité de problèmes en se basant sur la règle de décision de Bayes et ses dérivées (fonctions discriminantes linéaires et quadratiques), estimation de densités, problème de dimensions et l'estimation d'erreur, mais dès le début des années 80 [1][3], elle a connu une croissance rapide aidée par l'avènement de l'informatique et l'interaction entre plusieurs disciplines telles que la statistique, les mathématiques, les réseaux de neurones et la biologie. Ces arguments ont contribué à intégrer de nouvelles idées et à implémenter des algorithmes d'apprentissage complexes et de traiter des données de volumes importants jusqu'à considérer la RdF comme une branche de l'informatique tandis qu'elle était une branche des mathématiques.

I.2 Approches de la reconnaissance de formes

Nous allons résumer les quatre meilleures approches de la RdF, elles ne sont pas nécessairement indépendantes [1].

I.2.1 Appariement de gabarit (Template matching)

Cette approche est l'une des premières approches de la RdF, elle est basée sur la similitude entre deux quantités de même type (points, courbes, formes). Pour ce faire, la forme à reconnaître est comparée à un prototype ou un modèle (template) stocké auparavant. Souvent ce modèle est appris à partir d'un ensemble d'apprentissage et la mesure de similarité utilisée est la corrélation. Cette méthode, efficace dans certains domaines, souffre des limites si les formes présentent des variations dues au prétraitement, au changement du point de vue et aux fortes variations intra classes.

I.2.2 Approche statistique (Statistical approach)

Dans cette approche, chaque exemple est représenté par un vecteur caractéristique formé de d attributs (features, measurements) et est vu comme un point dans un espace de d dimensions. Son objectif est de trouver les frontières de décision dans cet espace caractéristique pouvant séparer les exemples en différentes classes. Dans la théorie de la décision statistique, les frontières de décision sont déterminées par les densités de probabilité des formes appartenant à chaque classe, ces densités peuvent être spécifiées ou apprises. On peut aussi utiliser, pour la classification, l'approche basée sur l'analyse discriminante: la forme de la frontière de décision (linéaire ou quadratique) doit être spécifiée et le choix d'une

meilleure frontière se fait par la classification des exemples d'apprentissage selon un certain critère tel que l'erreur quadratique moyenne.

I.2.3 Approche syntaxique (Syntactic approach)

Pour classifier des exemples complexes, on adopte la perspective hiérarchique; les exemples sont formés de simples sous exemples qui eux même composés de sous exemples très simplifiés appelés « primitives », donc il y a une analogie entre cette structure et la syntaxe du langage: les primitives remplacent l'alphabet et les exemples les phrases. Donc une grande collection d'exemples complexes peut être décrite par un petit nombre de primitives et de règles grammaticales inférées à partir des exemples d'apprentissage disponibles. Cette approche donne, en plus de la classification, une description de la construction des exemples à partir des primitives, cependant son implémentation souffre de la segmentation des exemples bruités (pour détecter les primitives) et de l'inférence de la grammaire à partir des données d'apprentissage, en plus, elle nécessite un grand ensemble d'apprentissage et un grand effort de calcul.

I.2.4 Réseaux de neurones (Neural networks)

Les réseaux de neurones sont des unités de calcul massivement interconnectées qui fonctionnent en parallèle. Ils présentent des capacités surprenantes d'apprentissage, d'auto adaptabilité, de généralisation et de tolérance aux fautes grâce à une procédure d'entraînement et d'apprentissage durant laquelle, le réseau adapte itérativement ses poids. Ces réseaux sont très attractifs pour réaliser des tâches de classification et d'extraction de caractéristiques (le perceptron multicouche et le réseau de fonctions à base radiales) et de clustering (réseau de Kohonen), ceci est dû au fait qu'ils [1] nécessitent peu de connaissances sur le domaine d'application et à la disponibilité d'algorithmes efficaces.

La majorité des réseaux de neurones populaires présente des similitudes avec les méthodes classiques de la RdF statistique et outre cette caractéristique, ils sont des approches unifiées pour l'extraction de caractéristiques et la classification et des procédures souples pour trouver les meilleures solutions non linéaires.

RdF statistique	Réseaux de neurones
Fonction discriminante linéaire	Perceptron
Analyse en composantes principales (ACP)	Réseau auto associatif et divers réseaux ACP
Estimation de probabilité a posteriori	Perceptron multicouche
Analyse discriminante non linéaire	Perceptron multicouche
Classificateur basé sur la densité de la fenêtre de Parzen	Réseau de fonctions à base radiales
Règle du PPV	LVQ de Kohonen

Tableau I.1. Relations entre les réseaux de neurones et les méthodes statistiques [1].

I.3 Systèmes de reconnaissance de formes

Pour assurer une tâche de reconnaissance de formes, un système de RdF comporte les étapes suivantes:

- Acquisition et prétraitement de données,
- Génération de caractéristiques,
- Représentation de données: extraction /sélection des caractéristiques,
- Classification (décision),

- Evaluation du système.

et opère en deux modes ; l'apprentissage et le test selon le modèle suivant [1]:

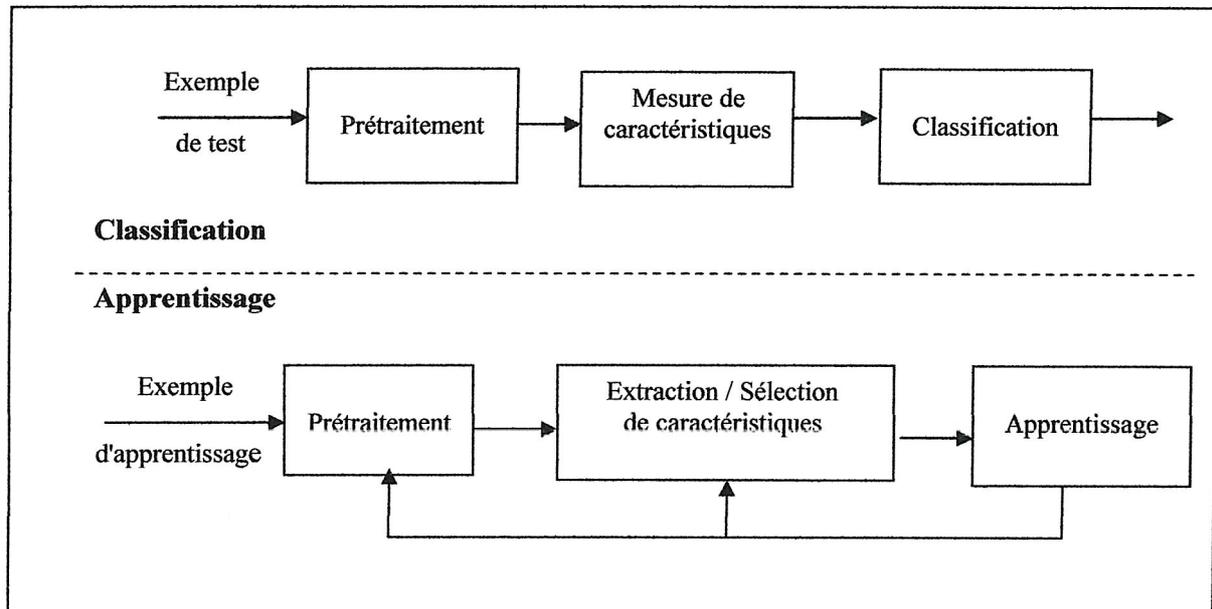


Figure I.1. Modèle pour la RdF statistique.

I.3.1 Acquisition et prétraitement de données

L'acquisition est assurée par des capteurs spécifiques (caméras, microphones, tablettes à digitaliser,...) qui fournissent des données de natures différentes dépendant du domaine d'application (visage humain, son de parole,...) et le prétraitement a pour but la segmentation des exemples, l'élimination du bruit et la normalisation des données.

I.3.2 Génération de caractéristiques

Elle est effectuée en collaboration avec un expert du domaine traité pour préserver les caractéristiques pertinentes et éliminer toute redondance.

I.3.3 Extraction de caractéristiques et réduction de dimensions

Cette étape a pour but la réduction de dimensions de l'espace caractéristique (partir d'un espace caractéristique de M dimensions pour obtenir un autre espace réduit de dimension $d < M$). Pour ce faire, on effectue soit la sélection qui consiste à chercher un sous ensemble d'attributs optimal de d dimensions soit l'extraction, qui vise à transformer ou à combiner les caractéristiques pour avoir un nouvel ensemble d'attributs de d dimensions.

I.4 Différentes approches de la reconnaissance de formes statistique

Selon la disponibilité d'informations sur la densité de probabilité conditionnelle d'appartenance aux classes, on peut regrouper les approches de la RdF selon le schéma de la figure I.2 [1].

- Si ces informations sont totalement connues, le classificateur peut être conçu en utilisant la règle de décision optimale de Bayes.
- Si la forme de cette densité est connue (ex: mélange de gaussiennes) mais seulement quelques paramètres sont inconnus (vecteur moyen, matrice de covariance) on aura un problème de décision paramétrique et ces paramètres seront remplacés par leurs valeurs estimées.
- Si la forme de la densité est inconnue, on est dans le mode non paramétrique et on peut l'estimer soit par la fenêtre de Parzen par exemple, soit en construisant directement la frontière de décision à partir de données d'apprentissage à l'aide de la règle du K plus proches voisins (KPPV) par exemple.

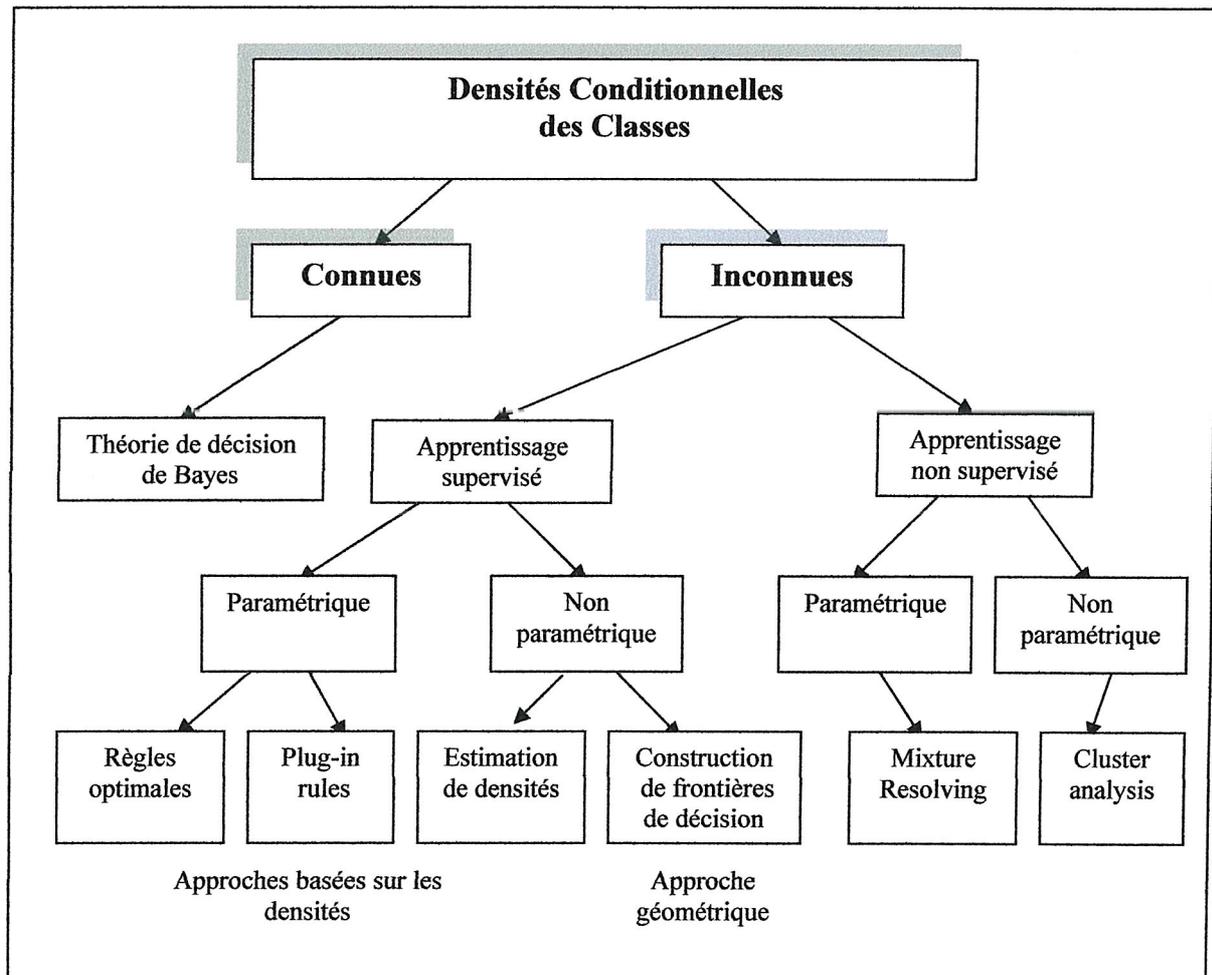


Figure I.2. Différentes approches de la RdF statistique [1].

II Classification

II.1 Représentation de données

On appelle **attribut** une variable réelle mesurable ou sensée décrivant un objet telle que la taille, la couleur, etc. Une **caractéristique** est soit un attribut ou une fonction d'un ou de plusieurs attributs. Les caractéristiques sont choisies de manière à ce qu'elles soient: **i)** indépendantes statistiquement, **ii)** semblables pour les formes d'une même classe et **iii)** dissemblables pour des formes de classes différentes.

Pour illustrer la représentation de données dans un espace caractéristique, nous avons représenté les trois classes de la base de données IRIS de Fisher [2] (voir l'annexe A). La

figure I.3(a) est un nuage de points dans un espace bidimensionnel où chaque point représente un exemple $X^{(q)} = [x_1^{(q)} x_2^{(q)}]$ formé de deux attributs seulement: x_1 représente le rapport longueur/ largeur des sépales et x_2 le rapport longueur/largeur des pétales de 150 fleurs formant cette base. Quant à la figure I.3 (b), elle représente les trois classes dans un espace 3D pour lequel, nous avons préservé la longueur, la largeur des sépales et la largeur des pétales.

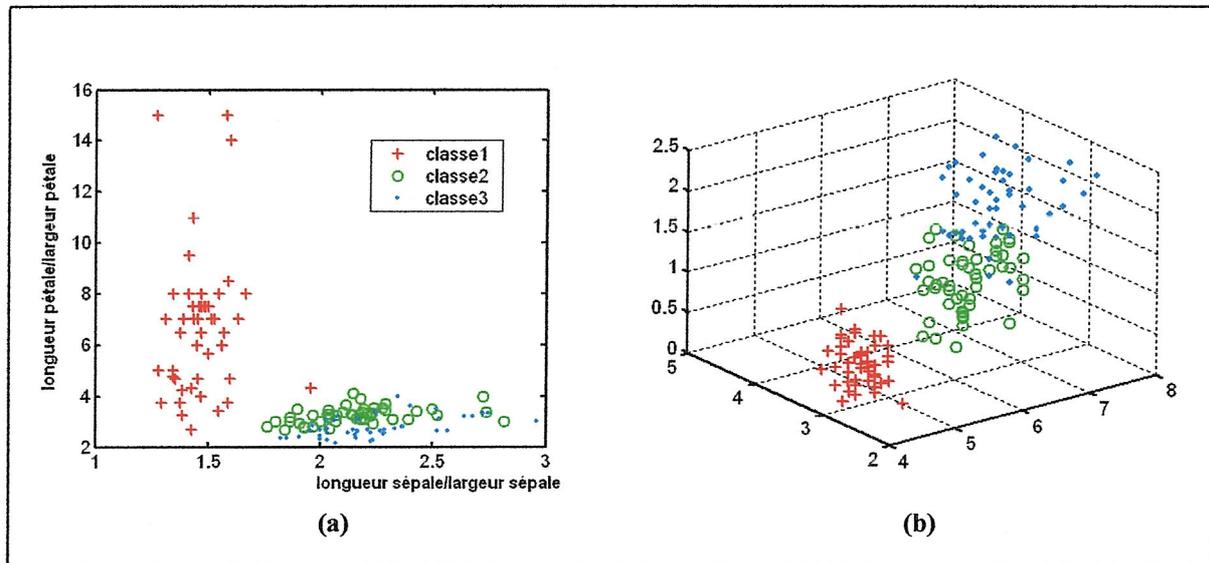


Figure I.3. Représentation des exemples de la base de données IRIS dans un espace :
(a) bidimensionnel et (b) tridimensionnel.

II.2 Classification et classificateur

Soit une population de Q objets où chacun est représenté au moyen d'un vecteur caractéristique $X^{(q)} = [x_1^{(q)} x_2^{(q)} \dots x_d^{(q)}]$ formé de d attributs et est vu comme un point dans l'espace caractéristique de d dimensions. Classifier ces objets revient à partitionner automatiquement le nuage de points $X^{(q)}$ en J classes (clusters) (C_1, C_2, \dots, C_J) de telle façon que les objets de la même classe soient les plus similaires possibles et les objets de classes différentes soient les plus dissemblables (voir figure I.4).

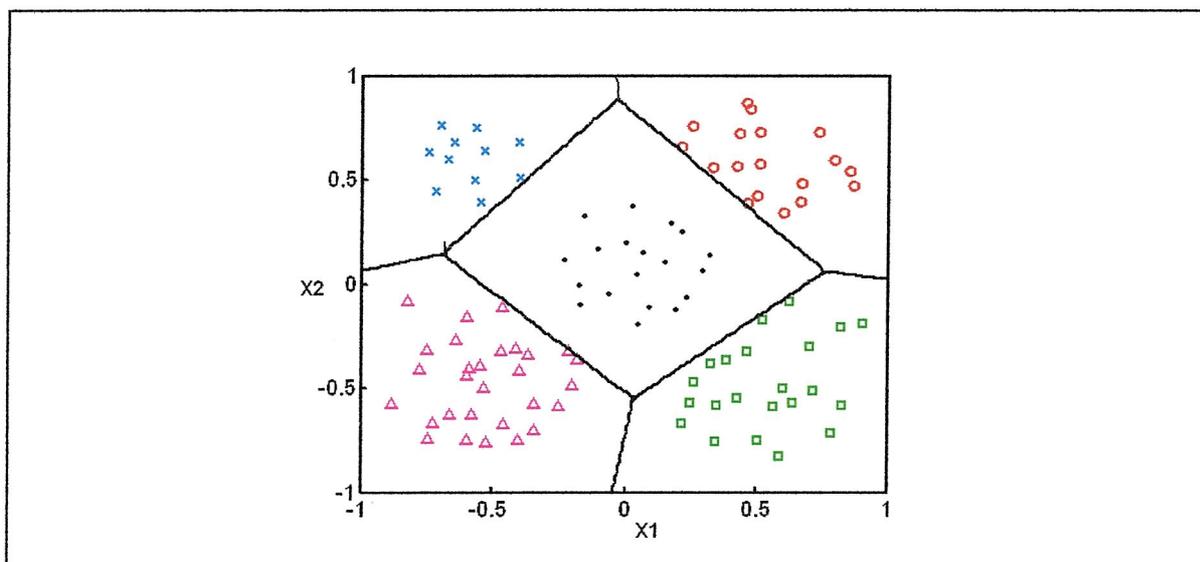


Figure I.4. Regroupement d'exemples en 5 classes.

La classification est effectuée par un classificateur qui renvoie l'appartenance d'un objet à classifier $X^{(q)}$ à une classe C_j parmi l'ensemble fini de J classes.

Ces classes peuvent être, selon l'application, visage ou non visage (détection de visages), acceptation ou rejet d'accès (reconnaissance de visages), type de gènes (bioinformatique), type de textes (classification de documents),...

Pour construire un classificateur, trois approches sont disponibles [1]:

1. l'approche la plus simple est basée sur la similitude entre les exemples à classifier c'est-à-dire tous les exemples similaires sont regroupés dans la même classe,
2. l'approche probabiliste basée sur la règle optimale de Bayes qui consiste à assigner l'exemple à classifier à la classe présentant la plus grande probabilité a posteriori,
3. l'approche géométrique qui consiste à déterminer les frontières de décision directement en optimisant un critère d'erreur.

II.2.1 Classification non supervisée (Unsupervised Classification)

Dans ce cas, on ne connaît pas a priori l'appartenance des exemples aux classes et le nombre de classes J peut être connu ou non. Une centaine d'algorithmes de clustering existe et d'autres continuent à apparaître [1] dont la majorité utilise deux techniques à savoir « iterative square error partitionnal » et « agglomerative hierarchical clustering », cette dernière fournit une suite de partitions en classes de plus en plus vaste (un arbre ou un dendogramme), quant à la première technique, elle vise à minimiser la variance intra-classes et à maximiser la variance inter-classes en adoptant un critère de clustering local ou global qui consiste pour le premier cas à affecter les exemples à leur classes en se basant sur leurs structures locales (ex: assigner un objet et ses K plus proches voisins à la même classe) tandis que pour le deuxième cas (ex: l'erreur quadratique) chaque classe est représentée par un prototype et l'exemple est assigné à la classe dont le prototype est le plus similaire. L'erreur totale à minimiser est donnée par:

$$E_j^2 = \sum_{j=1}^J e_j^2 \quad (\text{I.1})$$

Où J : nombre de classes

e_j^2 : erreur quadratique ou la variation intra-classes donnée par:

$$e_j = \sum_{q=1}^{n_j} (X_j^{(q)} - m_j)^T (X_j^{(q)} - m_j) \quad (\text{I.2})$$

avec: m_j vecteur moyen (centre de la $j^{\text{ème}}$ classe) défini par:

$$m_j = \left(\frac{1}{n_j}\right) \sum_{q=1}^{n_j} X_j^{(q)} \quad (\text{I.3})$$

et n_j le nombre d'exemples de la $j^{\text{ème}}$ classe tel que : $\sum_{j=1}^J n_j = Q$.

Parmi les algorithmes de clustering les plus connus, on compte: K-means, Fuzzy K-means, Mixture Decomposition,...

II.2.2 Classification supervisée (Supervised Classification)

Le nombre de classes est connu et on dispose d'un ensemble d'exemples prototypes qu'on sache leur classes a priori; pour ce type de classification, la séparation s'effectue soit en assignant à l'exemple à classifier, la classe représentée par les prototypes dont il ressemble le plus, soit en réalisant à partir d'un apprentissage supervisé, un modèle capable d'effectuer la classification.

II.3 Performances d'un classificateur

Le classificateur est, en premier lieu, conçu en utilisant les exemples d'apprentissage (phase d'apprentissage) puis évalué selon ses performances de classification sur un ensemble de test (phase de test). L'optimisation de ses performances sur l'ensemble d'apprentissage ne conduit pas forcément aux performances désirées sur l'ensemble de test, donc un classificateur qui généralise bien arrive à classifier correctement les exemples de test et la dégradation de ce pouvoir dépend de l'un de ces facteurs:

- la grande taille des vecteurs caractéristiques par rapport au nombre d'exemples d'apprentissage disponibles (ex: reconnaissance de visages),
- la petite taille de l'ensemble de test,
- le grand nombre des paramètres inconnus du classificateur (classificateurs polynomiaux, réseau de neurones de taille importante),
- le surapprentissage,
- la complexité du classificateur.

La performance d'un classificateur est mesurée en termes [1] :

- de taux d'erreur (erreur de classification),
- de taux de fausses acceptations (False Acceptance Rate),
- de taux de faux rejets (False Reject Rate),
- de taux de rejet,
- de probabilité d'erreur,
- de coût de mesure des caractéristiques,
- de temps nécessaire pour la décision.

Plusieurs méthodes sont utilisées pour estimer le taux d'erreur. Elles diffèrent selon la façon de partitionner l'ensemble de données en base d'apprentissage et base de test. Parmi lesquelles:

1. Leave-one-out : un classifieur est réalisé en utilisant $(Q - 1)$ exemples pour l'apprentissage et l'exemple restant pour le test. Cette opération est répétée Q fois avec différents ensembles de tailles $(Q - 1)$ ce qui est coûteux en temps de calcul. Cette technique a été utilisée par [52][57] pour estimer les résultats de leurs systèmes de reconnaissance de visages.

2. Hold out : il consiste à utiliser la moitié de l'ensemble de données pour l'apprentissage et l'autre pour le test [1][10].

3. Validation croisée (n fold cross validation) : la validation croisée est un compromis entre les méthodes Leave-one-out et le Hold out. Elle consiste à départager l'ensemble de données E en P ($1 \leq P \leq Q$) sous ensembles disjoints, en général 10, tels que:

$$E = E_1 \cup E_2 \cup \dots \cup E_p \quad \text{et} \quad E_i \cap E_j = \emptyset \quad \forall i \neq j$$

et utiliser $(P - 1)$ sous ensembles pour l'apprentissage et le sous ensemble restant pour le test. Cette opération est répétée P fois avant de moyenner les résultats. Moghaddam a appliqué une validation croisée d'ordre 5 pour estimer ses résultats [61][53].

4. Resubstitution method : tout l'ensemble de données disponibles est utilisé à la fois pour l'apprentissage et le test [1].

D'autres techniques existent [5] telles qu'apprendre sur les 10 premiers exemples et tester sur le 11^{ème}, apprendre sur les 11 premiers exemples et tester sur le 12^{ème} et ainsi de suite, puis moyenner les erreurs obtenues à chaque fois. Pour le choix d'un estimateur, [5] a adopté une règle heuristique :

- plus de 100 exemples, utiliser la validation croisée avec $P = 10$ ou le Hold out pour une très grande base,
- entre 50 et 100 exemples, utiliser le Leave-one-out.

II.4 Combinaison de classificateurs

La combinaison de plusieurs classificateurs améliore la performance de la classification en mettant au profit les performances de chaque classificateur pris seul. Elle est nécessaire si [1]:

1. certaines applications nécessitent l'utilisation de plusieurs classificateurs chacun développé dans un contexte et pour une représentation différente (ex: identification de personnes en fusionnant la voix, le visage et les empreintes digitales),
2. on dispose de plusieurs ensembles d'apprentissage obtenus dans des endroits et dans des périodes différentes,
3. plusieurs classificateurs ne possèdent pas les mêmes performances pour les mêmes exemples d'apprentissage,
4. la performance de certains classificateurs dépend de leurs paramètres internes (ex : poids initiaux dans le cas d'un réseau de neurones), d'où la combinaison pour bénéficier des performances de chacun.

II.5 Théorie de la décision

La théorie de la décision constitue une approche statistique fondamentale pour résoudre le problème de classification [6]. Parmi les règles de décision les plus répandues qui fournissent des frontières discriminantes, la règle de décision de Bayes, la règle du maximum de vraisemblance (un cas particulier de la règle de Bayes) et la règle de Neyman-Pearson.

II.5.1 La règle de décision optimale de Bayes

Cette règle a été développée par le mathématicien britannique Thomas Bayes au 18^{ème} siècle, elle peut se formuler comme suit [1][4]:

Pour un objet à classifier représenté par son vecteur caractéristique $X^{(q)} = [x_1^{(q)} x_2^{(q)} \dots x_d^{(q)}]$ de d dimensions, on cherche à l'assigner à une classe C_j parmi les J classes présentes sachant $P(C_j)$ la probabilité a priori d'avoir la classe C_j et $P\left(\frac{X}{C_j}\right)$ la probabilité conditionnelle d'observer X sachant la classe C_j . Si ces probabilités sont inconnues, on peut les estimer [4] à partir des exemples d'apprentissage disponibles.

La probabilité a posteriori d'appartenance de l'objet X aux classes C_j avec $j = 1, \dots, J$ est donnée par la règle de Bayes suivante:

$$P\left(\frac{C_j}{X}\right) = \frac{P\left(\frac{X}{C_j}\right) \cdot P(C_j)}{\sum_{j=1}^J P\left(\frac{X}{C_j}\right) \cdot P(C_j)} \quad (\text{I.4})$$

Selon «Maximum A Posteriori rule (MAP)» [1][4], X est assigné à la classe C_j si:

$$P\left(\frac{C_j}{X}\right) > P\left(\frac{C_i}{X}\right) \quad \text{pour tout } j \neq i \quad (\text{I.5})$$

Dans le cas d'égalité [4], l'objet peut être affecté indifféremment à l'une des classes.

Le classificateur de Bayes a été utilisé à des fins de reconnaissance par Brunelli et Poggio [74] et pour la détection de visages par Schneiderman et Kanade [73].

II.5.2 Erreur de classification (probabilité d'erreur)

Considérons un problème à deux classes pour lequel une surface de décision sépare l'espace des caractéristiques en deux domaines de décision D_1 et D_2 . En utilisant ces domaines pour classer des observations, on commet deux types d'erreur [6]:

- Une observation X peut être située dans un domaine D_1 alors qu'elle appartient effectivement à la classe C_2 ,
- Une observation X peut être située dans un domaine D_2 alors qu'elle appartient à la classe C_1

La probabilité d'erreur associée au partage d'un ensemble d'observations en deux classes selon leur position vis-à-vis des domaines D_1 et D_2 est alors:

$$\begin{aligned} S = P(\text{erreur}) &= P(X \in D_1, C_2) + P(X \in D_2, C_1) \\ &= P(\text{erreur}) = P(X \in D_1 | C_2)P(C_2) + P(X \in D_2 | C_1)P(C_1) \\ &= P(\text{erreur}) = \int_{D_1} P(X|C_2)P(C_2)dx + \int_{D_2} P(X|C_1)P(C_1)dx \end{aligned} \quad (\text{I.6})$$

La figure I.5 illustre la probabilité d'erreur pour un problème unidimensionnel

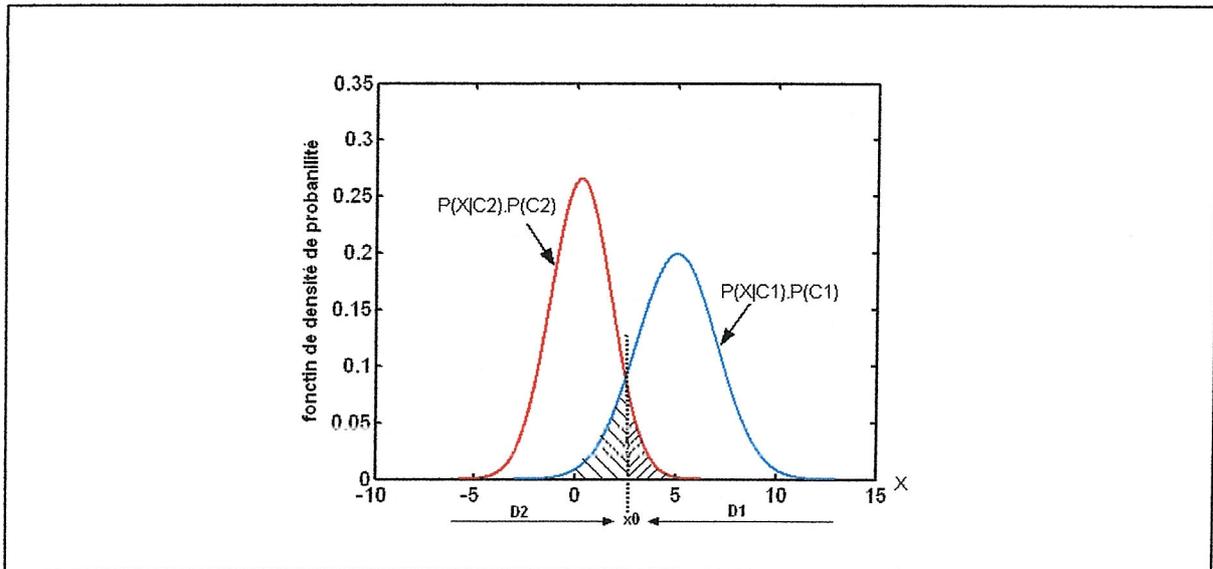


Figure I.5. Probabilité d'erreur pour un problème à une dimension et à deux classes.

Si les domaines D_1 et D_2 sont choisis arbitrairement, la probabilité d'erreur n'est pas minimale. Cette probabilité est minimale lorsque le point frontière entre D_1 et D_2 est située en x_0 . Dans ce cas la procédure revient à adopter la règle de décision suivante:

$$\text{décider } C_1 \quad \text{si} \quad P\left(\frac{X}{C_1}\right).P(C_1) > P\left(\frac{X}{C_2}\right).P(C_2) \quad (\text{I.7.a})$$

$$\text{décider } C_2 \quad \text{si} \quad P\left(\frac{X}{C_2}\right).P(C_2) > P\left(\frac{X}{C_1}\right).P(C_1) \quad (\text{I.7.b})$$

Le calcul précédent de la probabilité d'erreur peut se généraliser à un problème de J classes, en calculant, dans ce cas, la probabilité de prendre des décisions exactes au lieu de la probabilité de commettre des erreurs:

$$P(\text{exacte}) = \sum_{j=1}^J P(X \in D_j, C_j) \quad (\text{I.8.a})$$

$$P(\text{exacte}) = \sum_{j=1}^J P(X \in D_j | C_j).P(C_j) \quad (\text{I.8.b})$$

$$P(\text{exacte}) = \sum_{j=1}^J \int_{D_j} P(X|C_j).P(C_j) dx \quad (\text{I.8.c})$$

Cette stratégie est généralement citée comme référence pour apprécier les performances des procédures de classement et de classification. Le taux d'erreur qui lui est associé, en dessous duquel il est théoriquement impossible de descendre, est qualifié de **taux d'erreur optimal**; on parle alors de procédure de classement ou de classification optimales.

II.6 Le plus proche voisin (Nearest Neighbor)

Le plus proche voisin (PPV) est l'un des classificateurs non paramétriques les plus anciens et les plus utilisés en reconnaissance de formes jusqu'à nos jours. Son concept de base

a été introduit par Fix et Hodges [9] mais en 1967, Covert et Hart [11] ont donné une définition formelle de la règle du plus proche voisin et l'ont appliquée en RdF.

Ce classificateur requiert [8]: **i)** un ensemble de prototypes, **ii)** une règle de classement, et **iii)** une mesure de voisinage, qui pour un ensemble $E = \{X^{(1)}, X^{(2)}, \dots, X^{(Q)}\}$ de Q exemples appartenant à J classes attribue un nouvel exemple Y représenté par son vecteur caractéristique à la classes de l'exemple $X^{(q)}$ le plus proche à Y selon une distance choisie.

Parmi les distances adoptées: la distance quadratique, la distance de Manhalanobis, la distance de similarité cosinus (The cosine Similarity Metric) et la norme L_a définie, pour deux vecteurs $X = [x_1, x_2, \dots, x_d]$ et $Y = [y_1, y_2, \dots, y_d]$, par:

$$L_a = \left(\sum_{i=1}^d |x_i - y_i|^a \right)^{1/a} \quad (\text{I.9})$$

Pour $a = 2$, on obtient la distance euclidienne, l'instanciation la plus répandue de cette norme.

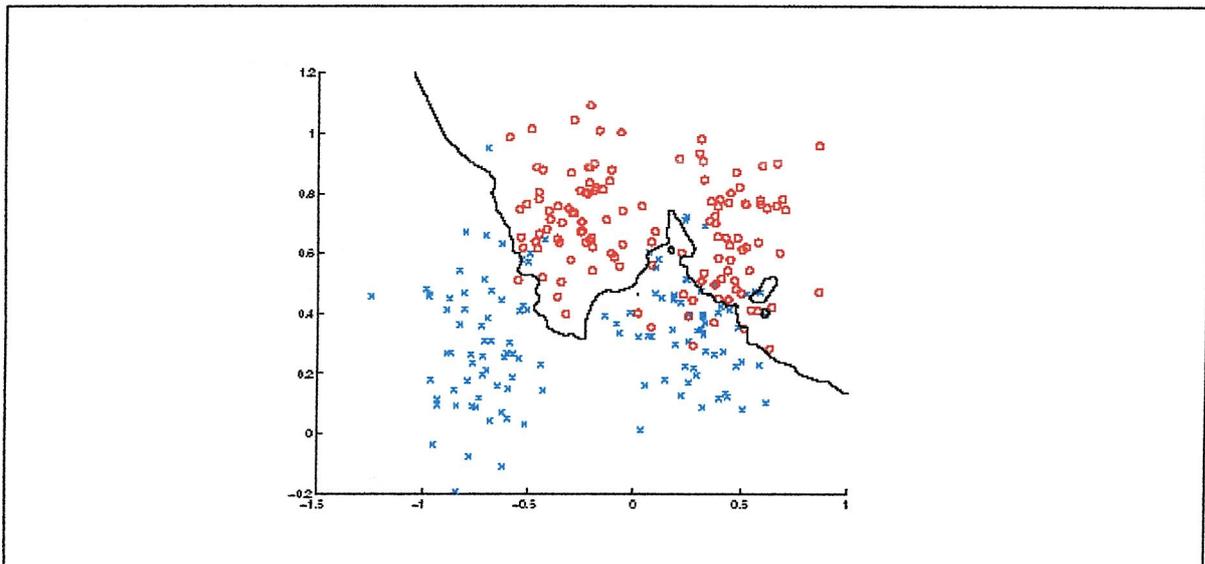


Figure I.6. Frontière de décision générée par le classificateur KPPV pour $K = 8$.

Une extension du classificateur PPV est la règle du KPPV dont le principe est d'affecter l'exemple à classifier à la classe majoritaire parmi les K plus proches voisins. Malgré que ce classificateur est simple et ne nécessite pas d'hypothèse sur les distributions de classes, il est gourmand en temps de calcul et possède une complexité de $O(Q \log k)$ pour traiter un exemple inconnu [8], c'est pourquoi le nombre de voisins « K » et la taille de l'ensemble de prototypes doivent être réduits.

II.7 Applications de la règle du plus proche voisin pour la reconnaissance de visages

Le PPV est largement utilisé en reconnaissance de visages (RdV) dès les premières approches [50][57][58] jusqu'au plus récentes [52][54][83][55][56], cependant nous devons

noter l'influence de la distance choisie sur la performance des systèmes de RdV comme l'a démontré Zhao et al.[58] en adoptant une distance pondérée, Bartlett et al.[84] et Drapper et al.[72] avec la similarité cosinus que les résultats obtenus avec ces mesures de similarité sont meilleurs qu'avec la distance euclidienne.

Récemment une variante de la distance euclidienne a été utilisée dans les systèmes de RdV tels que [52][83] où la distance est calculée entre deux matrices $A = [a_1 a_2 \dots a_n]$ et $B = [b_1 b_2 \dots b_n]$ comme suit:

$$d(A, B) = \sum_{i=1}^n \|a_i - b_i\|_2 \quad (\text{I.10})$$

où $\|a_i - b_i\|_2$ désigne la distance euclidienne entre les deux vecteurs colonnes a_i et b_i .

III Réduction de dimensions

Comme nous l'avons vu, pour classifier un objet, on doit le représenter par son vecteur caractéristique $X = [x_1 x_2 \dots x_d]$ de d dimensions, si cette dimension est trop grande, le classificateur conçu consomme aussi bien le temps de calcul que l'espace mémoire, ce qui peut dégrader ses performances. Ces problèmes, entre autres, sont regroupés sous le terme **malédiction de la dimension** (Curse of dimensionality) introduit par le mathématicien Bellman père de la programmation dynamique [3]. Donc il est nécessaire de réduire la dimension de l'espace caractéristique avant tout traitement sans toute fois dégrader le pouvoir discriminant et par conséquent le taux de classification. Pour ce faire, on fait appel aux méthodes d'extraction de caractéristiques (linéaires ou non linéaires).

III.1 Méthodes d'extraction de caractéristiques

Les méthodes d'extraction de caractéristiques visent la réduction de dimensions de l'espace caractéristique en projetant les données originales via une fonction de projection (linéaire ou non linéaire) dans un espace de dimension inférieure. Parmi les transformations linéaires:

- L'analyse en composantes principales (ACP),
- L'analyse discriminante linéaire (ADL),
- L'analyse factorielle (AF),
- L'analyse en composantes indépendantes (ACI),
- La poursuite évolutionnaire (PE),
- La poursuite de projection (PP).

et parmi les transformations non linéaires, on cite:

- L'analyse en composantes principales à noyau (Kernel PCA),
- L'analyse discriminante linéaire à noyau (Kernel LDA),
- Echelonnage multidimensionnel métrique (MDS),

Les réseaux de neurones, qui seront détaillés dans le chapitre suivant, peuvent être directement et efficacement utilisés pour l'extraction de caractéristiques dans le mode non supervisé. Parmi lesquels : la carte auto organisatrice, le perceptron multicouche [1][53] et le recirculation neural network [7]. Ils présentent l'avantage d'éviter le passage par le calcul de la matrice de covariance qui est très coûteuse à calculer.

Une extraction de caractéristiques linéaire (figure I.7 (a)) et une autre non linéaire (figure I.7 (b)) sont obtenues via les couches cachées des réseaux auto-associatif de la figure I.7. Pour le premier cas, l'architecture du réseau est d_3_d avec des fonctions d'activation linéaires. Les trois sorties de la couche cachée représentent les trois premières composantes principales.

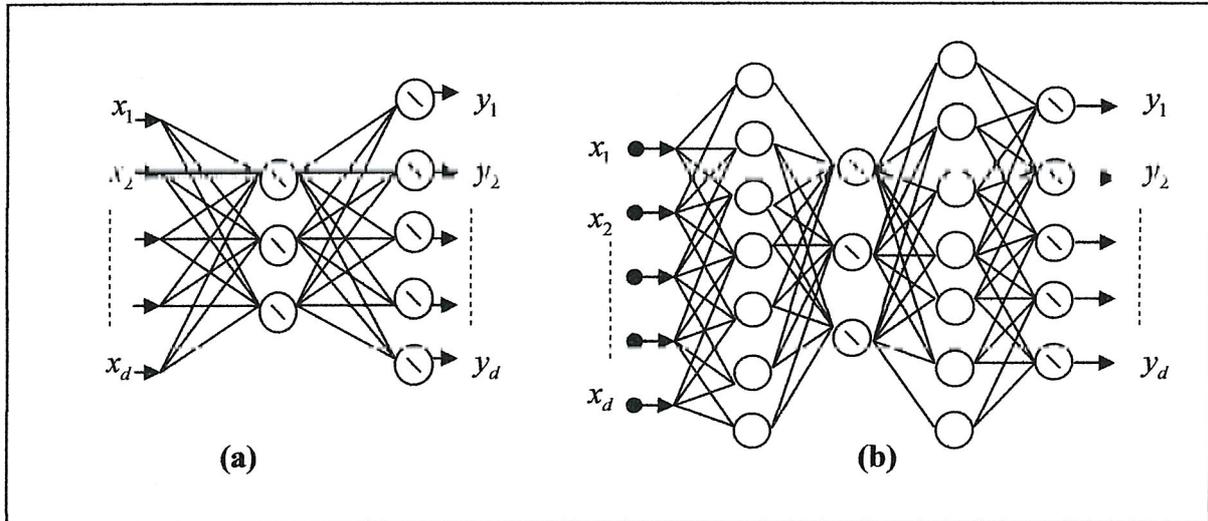


Figure I.7. Réseaux de neurones auto-associatifs pour l'extraction de caractéristiques :
(a) extraction linéaire et (b) extraction non linéaire

Pour le deuxième cas, on rajoute au réseau précédent deux autres couches cachées dont les fonctions d'activations sont des sigmoïdes. Ce réseau est dit [53] réseau de neurones multicouche auto codeur ACP non linéaire (NLPCA) puisque sa couche cachée intermédiaire (bottleneck layer) forme une représentation réduite avec une projection non linéaire donnant ainsi les trois composantes principales. Parmi toutes ces méthodes, nous allons nous focaliser sur deux méthodes très appliquées en RdV; l'analyse en composantes principales et l'analyse discriminante linéaire.

III.2 Analyse en composantes principales (ACP)

L'ACP est une technique linéaire d'extraction de caractéristiques proposée premièrement par Pearson en 1900 pour deux variables puis par Hotelling en 1933 pour un nombre quelconque de variables.

Sous l'hypothèse que les variables à traiter sont centrées (à valeur moyenne nulle), l'ACP devient similaire à la transformée de Karhunen-Loève (KLT) proposée indépendamment en 1947 par Karhunen et en 1955 par Loève [12] pour le traitement des signaux.

Pour une matrice $X(N \times M)$ dont les colonnes sont les observations d'un signal de N dimensions, KLT est une transformation orthogonale avec la base $\Phi = [\phi_1 \phi_2 \dots \phi_N]$ qui pour tout $M' \leq N$ minimise l'erreur donnée, pour le point X_i par:

$$\varepsilon(X_i) = \left\| X_i - \sum_{i=1}^{M'} (\phi_i^T X_i) \phi_i \right\| \quad (\text{I.11})$$

$$\text{avec } X = \begin{bmatrix} x_1^{(1)} \cdots x_M^{(1)} \\ \vdots \\ x_1^{(N)} \cdots x_M^{(N)} \end{bmatrix} = [X_1 X_2 \cdots X_M] \quad (\text{I.12})$$

La base Φ est obtenue en résolvant le problème de valeurs propres suivant:

$$\begin{cases} \Sigma = \frac{1}{M} \sum_{i=1}^M X_i X_i^T \\ \Lambda = \Phi^T \cdot \Sigma \cdot \Phi \end{cases} \quad (\text{I.13})$$

avec Λ : matrice diagonale contenant les valeurs propres de la matrice Σ ,
 Φ : matrice formée des vecteurs propres correspondant aux valeurs propres de Σ ordonnées en ordre décroissant.

III.2.1 Principe et algorithme

Sachant une population de N individus formant les lignes d'un tableau X et décrits par M variables numériques en colonnes ($M > 1$), on cherche une représentation de ces N individus dans un sous espace réduit de l'espace initial. Pour ce faire, on définit M' nouvelles variables combinaison linéaires des M variables originales tout en préservant le maximum possible d'informations. Ces M' sont dites composantes principales (C.P) et les axes qu'elles déterminent axes principaux.

L'algorithme suivant détaille les étapes de l'application de l'ACP :

Algorithme I.1: Analyse en Composantes Principales (ACP)

Entrées: Tableau individus/ Variables $X(N \times M)$

Sorties: Composantes Principales

1. Calculer la moyenne de X : \bar{X}
 2. Centrer les variables (retrancher à chaque variable sa moyenne): X_c
 3. Calculer la matrice de covariance des variables centrées : $C = \frac{1}{N} X_c^T X_c$
 4. Calculer les valeurs propres et les vecteurs propres de la matrice de covariance C
 5. Trier les valeurs propres en ordre décroissant: $\lambda_1 > \lambda_2 > \cdots > \lambda_M$
 - Préserver les M' premiers vecteurs propres correspondant à M' valeurs propres les plus grandes,
 - Rassembler les M' vecteurs propres dans la matrice $V = [v_1 v_2 \cdots v_{M'}]$
 6. Calculer les composantes principales Y : $Y = X_c \cdot V$
-

L'application de l'algorithme I.1, nous a donné la projection de la base de données IRIS dans un espace bidimensionnel. Avec deux composantes principales, l'ACP parvient à séparer les trois classes comme le montre la figure I.8.

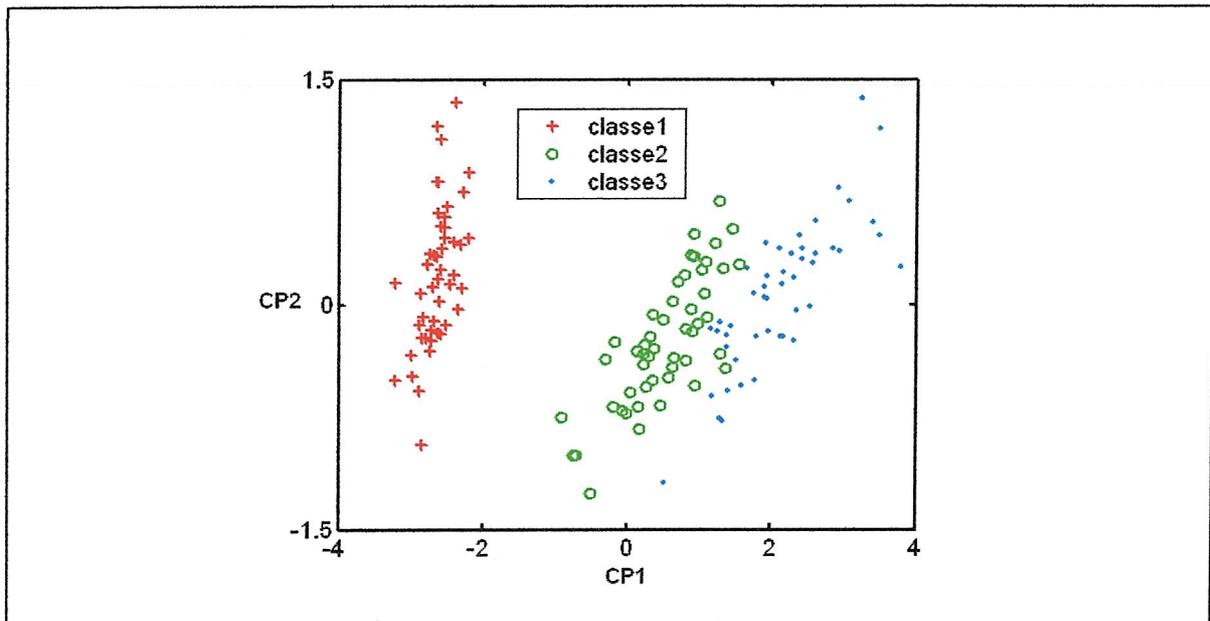


Figure I.8. Projection de la base IRIS par l'ACP dans un espace bidimensionnel.

III.2.2 Propriétés

L'application de l'ACP apporte plusieurs avantages :

- Représentation de données compacte,
- Composantes principales décorréliées et centrées,
- Base de l'espace de projection orthonormale: $v_k^T v_k = I$,
- Reconstruction de données approximative selon: $X \approx Y.V^T$,
- L'ACP donne des composantes principales qui sont les plus expressives [66].

III.2.3 Inconvénients

Les inconvénients de l'ACP sont dus essentiellement à la réduction de dimensions [10] :

- L'ACP fait perdre le sens réel des données (longueur, largeur, couleur,...) après la transformation en composantes principales,
- L'ACP dégrade le pouvoir discriminant à cause de la négligence de certaines informations contenues dans les vecteurs propres non retenus,
- L'ACP est une transformation linéaire qui ne reflète pas les non linéarités qui peuvent être dans les données,
- La matrice de covariance est coûteuse à calculer en grande dimension.

Pour remédier au 4^{ème} inconvénient, l'ACP est souvent appliquée à l'aide d'une décomposition en valeurs singulières (SVD) dont le principe est le suivant :

Pour une matrice $X(N \times M)$ avec $N \geq M$, la décomposition en valeurs singulières donne:

$$X = U \cdot D \cdot V^T \quad (\text{I.14})$$

avec $U(N \times M)$ et $V(M \times M)$ deux matrices dont les colonnes sont orthonormales,
 D : matrice diagonale contenant les valeurs singulières de X (racines carrées des valeurs propres de $X^T \cdot X$),
 V : matrice contenant les vecteurs propres de $X^T \cdot X$.

III.3 Analyse discriminante linéaire ADL (Linear Discriminant Analysis: LDA/ Fisher Discriminant Analysis: FDA)

La discrimination linéaire de Fisher est une technique classique en RdF, développée à l'origine par Robert Fisher en 1936 [2] pour la classification taxonomique. Mais son champ d'application recouvre, parmi d'autres, la reconnaissance de parole, la classification de textes, l'extraction des contours des lèvres [64], la reconnaissance des gestes des mains, la reconnaissance de visages [57][58] et la recherche des images (Image Retrieval) [66].

Similaire à l'ACP, l'ADL est une technique linéaire d'extraction de caractéristiques et de réduction de dimension, mais qui exige la connaissance a priori d'appartenance des exemples à leur classes (mode supervisé) et la disponibilité d'au moins deux exemples par classes. L'ADL, au contraire de l'ACP, fournit les caractéristiques les plus discriminantes [66].

III.3.1 Principe

Pour un problème de C classes, l'ADL donne une fonction discriminante linéaire vérifiant :

$$Z = f(X) = W^T \cdot X \quad (\text{I.15})$$

avec X : vecteur colonne représentant un exemple à classifier,
 W : matrice optimale de projection,
 Z : vecteur caractéristique.

en projetant les exemples à classifier dans un espace plus approprié pour une représentation discriminante de dimension maximale $(C - 1)$. Appliquer l'ADL revient à déterminer des axes orthogonaux qui maximisent le critère suivant :

$$J(W) = \frac{\|W^T S_b W\|}{\|W^T S_w W\|} \quad (\text{I.16})$$

signifiant qu'ils maximisent la variance inter-classes et minimisent la variance intra-classes où S_b et S_w désignent, respectivement, les matrices de covariance inter et intra classes définies par :

$$S_b = \sum_{c=1}^C n_c (\bar{X}_c - \bar{X})(\bar{X}_c - \bar{X})^T \quad (\text{I.17.a})$$

$$S_w = \sum_{c=1}^C \sum_{i=1}^{n_c} (X_i - \bar{X}_c)(X_i - \bar{X}_c)^T \quad (\text{I.17.b})$$

C : nombre de classes,
 n_c : nombre d'exemples appartenant à la classe c ,
 \bar{X}_c : vecteur moyen de la classe c ,

\bar{X} : vecteur moyen total.

Il a été démontré [66] que le critère $J(W)$ est maximal pour une matrice de projection W dont les colonnes sont les vecteurs propres associés à M' valeurs propres les plus grandes. Pour cela, on doit résoudre le problème suivant:

$$S_b w_i = \lambda_i S_w w_i \quad \text{avec} \quad i = 1, \dots, M' \quad (\text{I.18})$$

$$\text{d'où } W_{opt} = \arg \max \left| \frac{W^T S_b W}{W^T S_w W} \right| = [w_1 w_2 \dots w_{M'}] \quad (\text{I.19})$$

III.3.2 Calcul de la matrice de projection optimale

Il y a plusieurs techniques pour résoudre le problème de valeurs propres donné par l'équation (I.18); la méthode directe consiste à inverser la matrice de covariance S_w et calculer les valeurs propres de $(S_w^{-1} S_b)$, mais cette technique est numériquement instable et coûteuse en temps de calcul parce que la matrice $(S_w^{-1} S_b)$ est non symétrique, la matrice S_w est souvent singulière [66] et de grande dimension [59]. Le problème peut être donc résolu par une décomposition de la matrice S_w . Puisque cette dernière est réelle symétrique, il existe deux matrices: Q orthonormale et Λ_Q diagonale qui vérifient:

$$S_w = Q \Lambda_Q Q^T \quad (\text{I.20})$$

L'équation (I.18) devient:

$$\begin{aligned} S_b W &= S_w W \Lambda_w = Q \Lambda_Q Q^T W \Lambda_w = Q \Lambda_Q^{1/2} \Lambda_Q^{1/2} Q^T W \Lambda_w \\ (Q \Lambda_Q^{1/2})^{-1} S_b W &= \Lambda_Q^{1/2} Q^T W \Lambda_w \\ \Lambda_Q^{-1/2} Q^T S_b W &= \Lambda_Q^{1/2} Q^T W \Lambda_w \end{aligned} \quad (\text{I.21})$$

en posant $R = \Lambda_Q^{-1/2} Q^T$ et $W_R = \Lambda_Q^{1/2} Q^T W$, l'équation (I.21) devient:

$$\begin{aligned} R S_b W &= W_R \Lambda_w \\ R S_b R^T R W &= W_R \Lambda_w \\ (R S_b R^T) W_R &= W_R \Lambda_w \end{aligned} \quad (\text{I.22})$$

Cette dernière équation donne Λ_w et W_R et par conséquent la matrice W s'obtient comme suit:

$$W = Q \Lambda_Q^{-1/2} W_R \quad (\text{I.23})$$

III.3.3 Algorithme

L'algorithme suivant détaille les étapes permettant l'application de l'analyse discriminante linéaire :

Algorithme I.2: Analyse discriminante linéaire (ADL)**Entrées:** tableau individus/variables $X(N \times M)$ **Sortie:** données projetées

1. Calculer la moyenne de chaque classe: $\bar{X}_c = \frac{1}{n_c} \sum_{i=1}^{n_c} X_i$
2. Calculer la moyenne totale: $\bar{X} = \frac{1}{N} \sum_{j=1}^N X_j$ avec $N = \sum_{c=1}^C n_c$
3. Calculer la matrice de covariance intra-classe (nous proposons la forme matricielle): $S_w = B.B^T$ avec $B = [\phi_1^{(1)} \phi_2^{(1)} \dots \phi_{n_1}^{(1)} : \phi_1^{(2)} \phi_2^{(2)} \dots \phi_{n_2}^{(2)} : \dots \phi_1^{(C)} \phi_2^{(C)} \dots \phi_{n_C}^{(C)}]$ et $\phi_i^{(c)} = (X_i - \bar{X}_c)^T$ pour $i = 1, \dots, n_c, c = 1, \dots, C$
4. Calculer la matrice de covariance inter-classes (nous proposons la forme matricielle): $S_b = n_c.A.A^T$ avec $A = [\varphi_1 \varphi_2 \dots \varphi_C]$ et $\varphi_c = (\bar{X}_c - \bar{X})^T$
5. Calculer les valeurs et les vecteurs propres de la matrice $S_w^{-1}.S_b$: λ_k et w_k
6. Trier ces valeurs propres en ordre décroissant
7. Retenir M' vecteurs propres correspondant à M' valeurs propres les plus grandes.
Former une matrice dont les colonnes sont les vecteurs propres :
 $W = [w_1 w_2 \dots w_{M'}]$
8. Projeter les données d'entrées sur un sous espace : $Z = X.W$

Nous avons appliqué cet algorithme sur la base de données IRIS de Fisher où $C = 3$, $n_c = 50$, $N = 150$ et $M' = 2$ afin d'illustrer le pouvoir discriminant de l'analyse discriminante linéaire. La projection de cette base dans un espace bidimensionnel permet la séparation des trois classes mieux que celle par l'ACP.

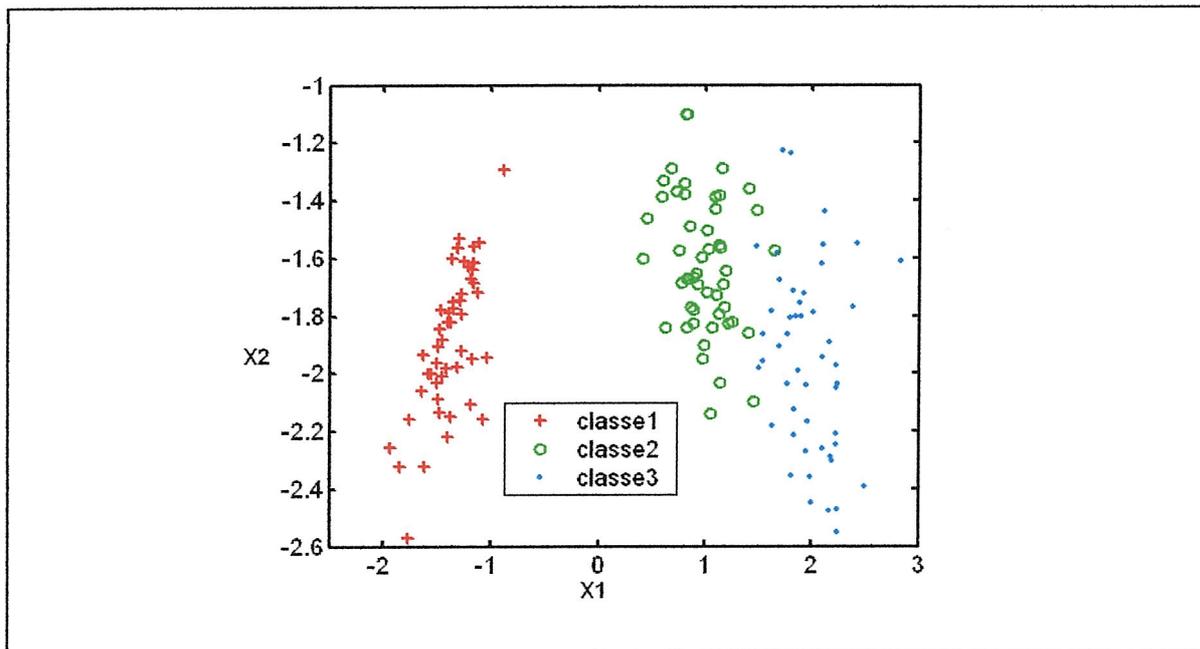


Figure I.9. Projection de la base IRIS par l'ADL dans un espace bidimensionnel.

III.4 Méthodes d'extraction de caractéristiques et de réduction de dimensions appliquées dans la reconnaissance de visages

La reconnaissance de visages (RdV) est un domaine où les méthodes de réduction de dimensions (linéaires et non linéaires) sont largement appliquées [50][85][76] et en particulier l'ACP qui est souvent combinée avec d'autres techniques telles que subspace LDA [57][58], l'ACI [84] et l'analyse discriminante de Fisher bidimensionnelle bilatérale (B2DFDA) [91]. Vu leur importance, de nombreuses contributions ont porté sur la comparaison de ces méthodes [53][68][72].

L'ACP et L'ADL restent les plus attractives; elles attirent de plus en plus les chercheurs qui ne cessent pas d'apporter des améliorations et/ou de proposer de nouvelles approches dérivées d'elles, soit dans l'espace monodimensionnel telles que Probabilistic PCA [71], Robust PCA [70] ou dans l'espace bidimensionnel telles que l'ACP2D [52], l'ACP2D bilatérale et l'ACP2D à noyau [83], l'ADL2D [54][77][87], l'ADL bilinéaire [55] et l'ADF2D à noyau [91].

Conclusion

Une étape très importante dans tout système de reconnaissance des formes est celle de l'extraction des caractéristiques et de réduction de dimensions. Un aperçu de plusieurs méthodes proposées récemment, pour la reconnaissance de visages, a été donné, mais en se focalisant sur l'analyse en composantes principales et l'analyse discriminante linéaire, nous avons pu présenter le principe et l'objectif de ces méthodes. Elles visent à préserver, à partir d'un volume important de données, les informations pertinentes qui discriminent au mieux les classes tout en réduisant la taille de l'espace caractéristique pour aboutir à un système de reconnaissance de formes performant en précision et en vitesse.

Parmi les classificateurs utilisés, nous intéressons aux réseaux de neurones qui seront étudiés dans le chapitre suivant.

Chapitre II

Réseaux de Neurones Artificiels

Introduction

Délaissés près de 20 ans après la publication du fameux livre du Marvin Minsky et Seymour Papert, en 1969, démontrant les limitations du perceptron qui ne parvient qu'à séparer les classes linéairement séparables, les réseaux de neurones ont vu encore une fois le jour vers le milieu des années quatre vingt dès la découverte d'un algorithme estimé puissant ; c'est l'algorithme de rétropropagation de l'erreur communément appelé back-propagation. Il est devenu populaire, grâce aux applications spectaculaires telles que le système NET-Talk qui apprend à lire à haute voix un texte anglais et le système WISARD pour la reconnaissance de visages.

Le pouvoir de généralisation, la capacité d'auto adaptabilité et le calcul parallèle ont fait des réseaux de neurones des classificateurs très attractifs dans divers domaines. En vu de les appliquer à des fins de reconnaissance de visages, ce chapitre, divisé en deux parties, rappelle, dans la première, les éléments de base des réseaux de neurones, et présente dans la deuxième, le perceptron multicouche (PMC), le type le plus populaire et étudié des réseaux de neurones. Son algorithme d'apprentissage, les techniques d'initialisation des poids, et le choix d'une topologie optimale seront également discutés.

I. Généralités sur les Réseaux de Neurones Artificiels

I.1 Neurone biologique

Un neurone biologique est une cellule nerveuse formant l'élément de base du système nerveux. Le cortex humain en contient environ 13 à 15 milliards arrangés dans une structure hiérarchique de 6 couches différentes, chaque neurone est connecté à environ 5000 autres neurones spécialisés dans la transmission, le traitement et le stockage d'informations. La figure II.1 schématise la structure du neurone biologique.

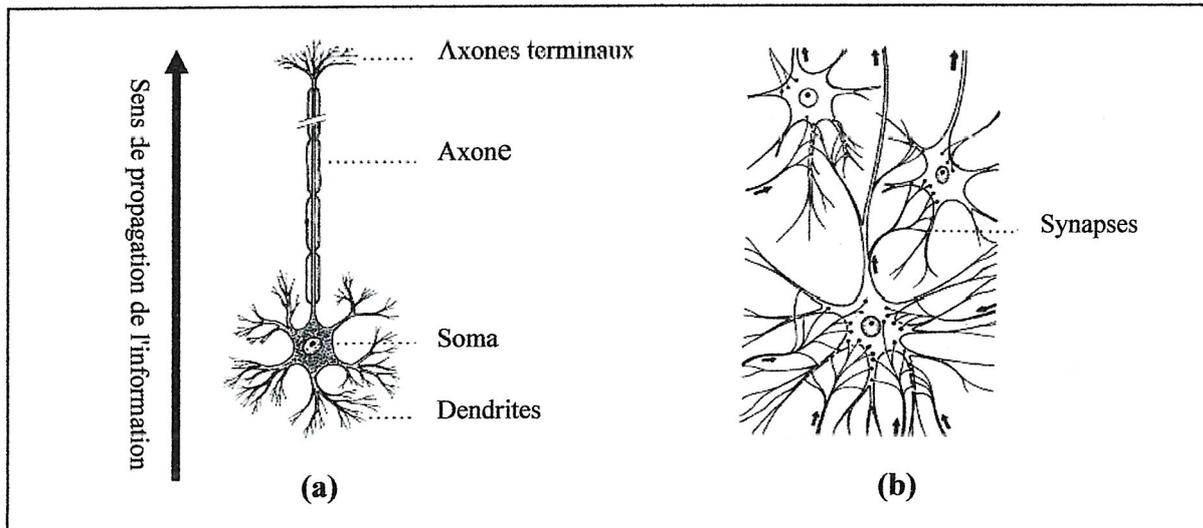


Figure II.1. Structure d'un neurone biologique:
 (a) Neurone biologique et (b) Réseau de neurones biologiques.

I.1.1 Structure du neurone biologique

Grossièrement, on peut diviser le neurone biologique en quatre parties:

- **les dendrites** : sont de fines extensions tubulaires de quelques dixièmes de microns de diamètre et d'une longueur de dizaines de microns formant l'entrée du neurone biologique qui capte l'information. Les dendrites se trouvent en grande quantité autour du soma, c'est pourquoi on parle souvent d'arborisation dendritique ou de chevelure dendritique.
- **le corps cellulaire (Soma)**: il contient le noyau cellulaire qui baigne dans un cytoplasme et c'est le centre d'influx nerveux. Sa forme est pyramidale ou sphérique dans la plupart des cas et est de quelques microns de diamètre. Il se ramifie pour former les dendrites.
- **l'axone (Axon)**: est une fibre nerveuse qui transmet l'influx nerveux aux autres neurones. Selon le type du neurone, la longueur de l'axone peut varier de quelques microns à 1.50m. A sa fin, une arborisation terminale où se trouve un nombre plus ou moins important de terminaisons synaptiques.
- **les synapses (Synapse)**: sont l'unique interface entre deux neurones assurant le transfert d'informations. Un neurone en contient de plusieurs centaines à une dizaine de milliers.

I.1.2 Fonctionnement du neurone biologique

Le soma et l'axone sont recouverts d'une membrane qui génère des impulsions électriques. La transmission de l'information est toujours dans le même sens: de l'axone du neurone amont vers les dendrites du neurone aval via les synapses. Pour assurer ce transfert; les dendrites captent les signaux électriques venant d'un ou de plusieurs neurones et le soma recueille et concentre l'ensemble des informations en effectuant une somme. Si cette somme dépasse un certain seuil alors il y a une émission d'un potentiel d'action à travers l'axone vers les dendrites des neurones cibles, sinon le neurone est passif et n'émet rien.

I.2 Neurone formel (Formal Neuron)

En s'inspirant du fonctionnement du neurone biologique, Warren McCulloch et Walter Pitts en 1943 [30], ont introduit pour la 1^{ère} fois un modèle mathématique très simplifié modélisant le neurone biologique; c'est le **neurone formel** schématisé en figure II.2. Ce dernier était discret et basé sur la logique binaire (fonctionnement tout ou rien) pour réaliser des fonctions logiques telles que le OU logique et le ET logique comme le montre la figure II.3. Cependant ce modèle ne possédait pas une règle d'apprentissage.

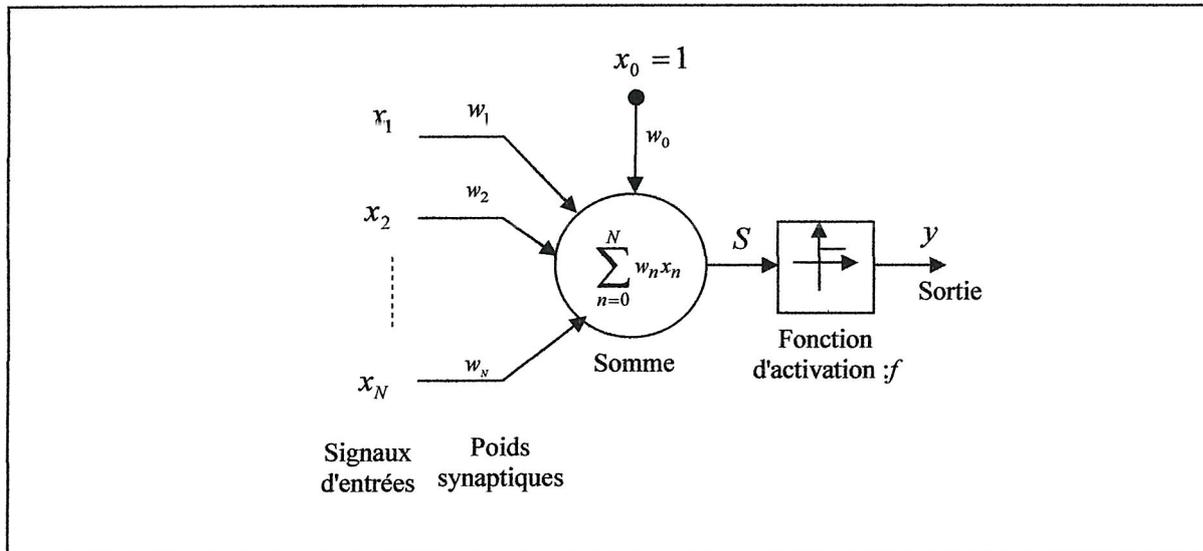


Figure II.2. Le neurone formel de McCulloch et Pitts.

Similaire au neurone biologique, le neurone formel possède N entrées, en général réelles : x_1, x_2, \dots, x_N associées à N poids synaptiques : w_1, w_2, \dots, w_N mesurant sa perméabilité. Le paramètre w_0 est dit biais, et par convention pour simplifier les formules mathématiques, il est associé à une entrée unitaire.

Ce neurone réalise une somme pondérée de ses entrées, il est actif si cette somme est positive et donne ainsi en sortie une réponse y modélisant l'impulsion électrique générée par l'axone, dans le cas contraire il est passif:

$$y = f(S) = f\left(\sum_{n=0}^N w_n x_n\right) = \begin{cases} 1 & \text{si } S \geq 0 \\ 0 & \text{si } S < 0 \end{cases} \quad (\text{II.1})$$

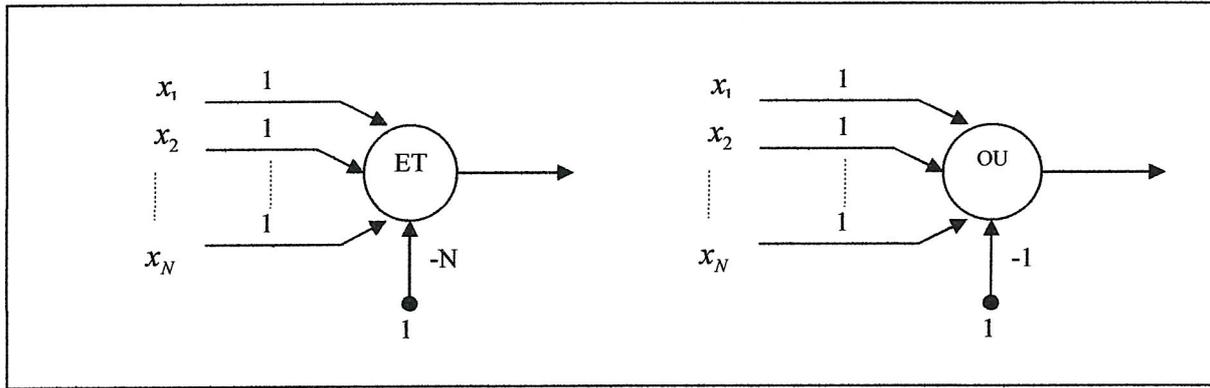


Figure II.3. Implémentation du ET logique et du OU logique à l'aide d'un neurone formel.

Interprétation géométrique du fonctionnement du neurone formel

Les N entrées du neurone peuvent être interprétées [15] comme les coordonnées d'un point $X = [x_1, x_2, \dots, x_N]$ dans l'espace euclidien E_N de N dimensions et le neurone formel à seuil, dans ce cas, partitionne cet espace en deux sous espaces H^+ et H^- délimités par un hyperplan (une ligne droite dans E_2 et un plan dans E_3). Les coefficients de cet hyperplan sont les poids synaptiques et son équation est donnée par la formule suivante:

$$w_0 + \sum_{n=1}^N w_n x_n = 0 \tag{II.2}$$

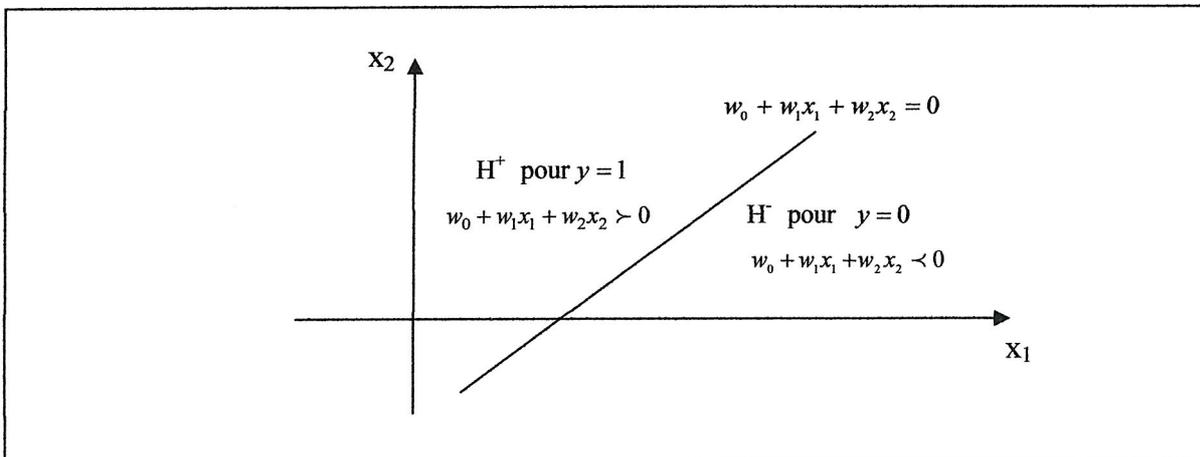


Figure II.4. Classification linéaire par un neurone formel dans le cas bidimensionnel.

Les points de l'espace caractéristique appartenant au demi-espace H^+ et ceux appartenant au demi-espace H^- vérifient respectivement les inégalités suivantes:

$$\text{si } y = 1 \quad \text{alors} \quad w_0 + \sum_{n=1}^N w_n x_n > 0 \tag{II.3.a}$$

$$\text{si } y = 0 \quad \text{alors} \quad w_0 + \sum_{n=1}^N w_n x_n < 0 \tag{II.3.b}$$

I.3 Perceptron

Conçu par Rosenblatt en 1958 [26], le perceptron est le premier réseau de neurones proprement dit doté d'un algorithme d'apprentissage. Il était destiné à la perception et la réalisation de certaines fonctions logiques avec une architecture simple (figure II.5) comportant une couche d'entrée dite, dans le modèle original, rétine (retina) et une couche de sortie dite couche d'association (association).

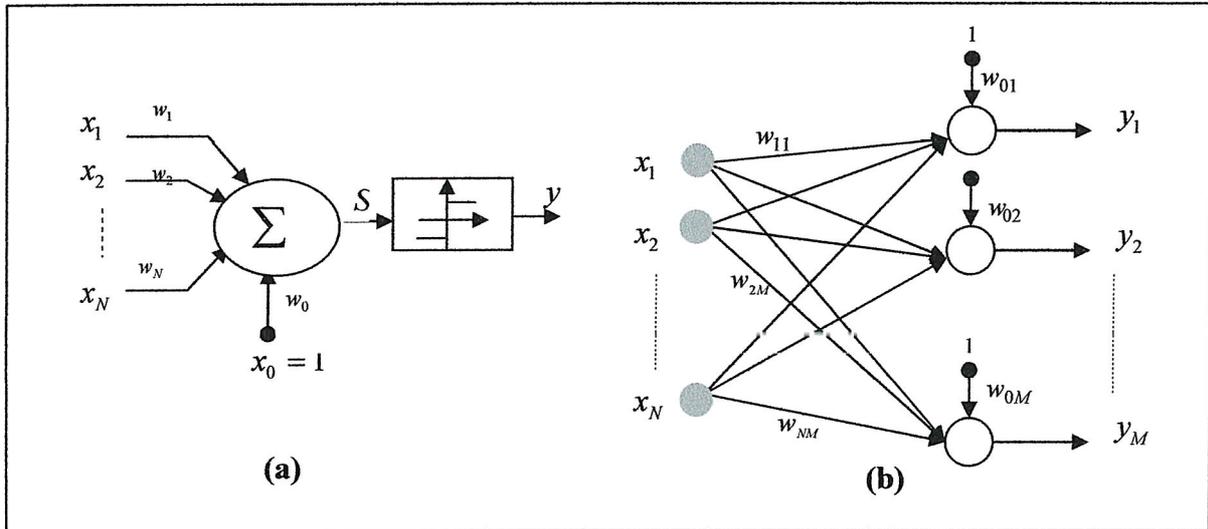


Figure II.5. Schématisation d'un perceptron: (a) Perceptron et (b) Réseau de perceptrons.

Les entrées du perceptron peuvent être booléennes ou réelles et la fonction d'activation à l'origine était la fonction Heaviside donnant en sortie une réponse dont les composantes sont binaires (0 ou 1) et ayant l'expression suivante:

$$y_m = f\left(\sum_{n=0}^N w_{nm}x_n\right) \quad \text{avec } m = 1, \dots, M \quad (\text{II.4})$$

w_{nm} : sont les poids synaptiques liant l'entrée n et le neurone m .

La règle d'apprentissage du perceptron, connue avec différentes appellations: la règle de Widrow-Hoff, la règle Delta ou tout simplement la règle d'apprentissage du perceptron (The Perceptron Learning Rule), consiste à ajuster itérativement ses poids synaptiques en corrigeant l'erreur entre la sortie cible (sortie désirée) et la sortie du réseau. Si l'on dispose d'un ensemble de Q exemples représentés par des vecteurs caractéristiques $X^{(q)} = [x_1^{(q)} x_2^{(q)} \dots x_N^{(q)}]$ associés à leur sorties désirées $d^{(q)} = [d_1^{(q)} d_2^{(q)} \dots d_M^{(q)}]$, alors l'adaptation des poids à l'itération $(i+1)$ se fait selon la règle:

$$w_{nm}^{(i+1)} = w_{nm}^{(i)} + \eta \cdot (d_m^{(q)} - y_m^{(q)}) \cdot X^{(q)} \quad \text{avec } n = 1, \dots, N \text{ et } m = 1, \dots, M \quad (\text{II.5})$$

où η est une petite constante positive ($0 < \eta < 1$), dite pas ou taux d'apprentissage, sur laquelle dépend la convergence de cet algorithme d'apprentissage.

En examinant l'équation (II.5), on peut dire que le perceptron n'apprend que s'il commet des erreurs; c'est-à-dire si $d^{(a)}$ diffère de $y^{(a)}$ il ajustera ses poids sinon il n'y aura pas d'ajustement et la procédure d'apprentissage se termine lorsque le perceptron ne commet plus d'erreurs [25].

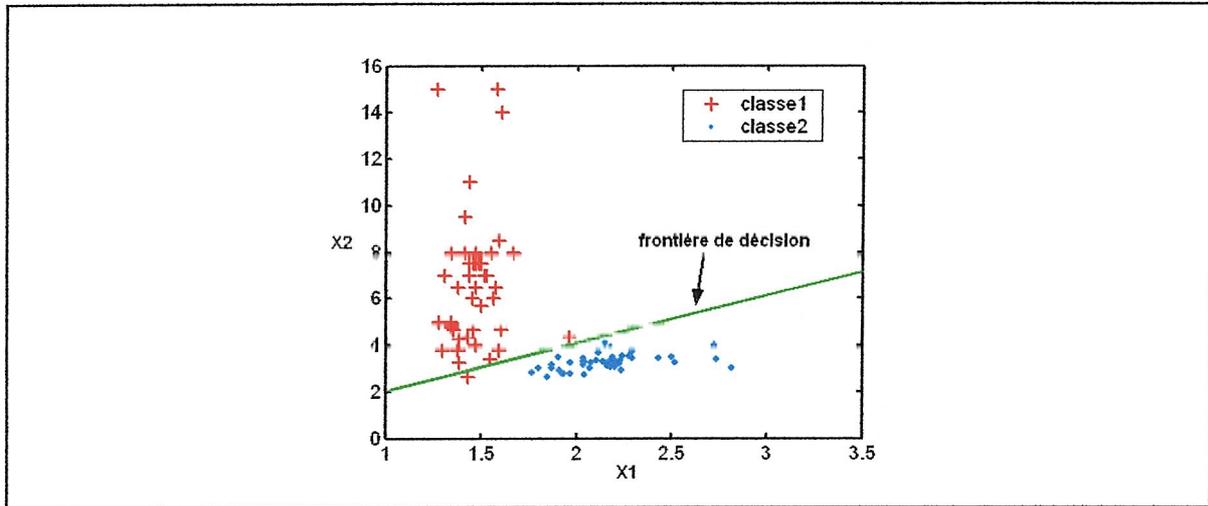


Figure II.6. Frontière de décision générée par un perceptron séparant deux classes de la base IRIS.

La figure II.6 illustre la partition linéaire de l'espace caractéristique bidimensionnel en deux sous espaces contenant deux classes (Sétosa et Versicolor) de la base Iris (voir l'annexe A), que nous l'avons obtenue par l'entraînement d'un perceptron (figure II.5 (a)) ayant deux entrées, une seule sortie, la fonction signe comme fonction d'activation et un taux d'apprentissage $\eta = 0.3$.

Limitation du perceptron

D'après le théorème de convergence du perceptron [26], son algorithme d'apprentissage converge vers une solution optimale (vecteur optimal de poids), si elle existe, après un nombre fini d'itérations si les exemples présents en entrée du perceptron appartiennent à des classes linéairement séparables. Cependant, Minsky et Papert en 1969 [32] à travers l'exemple du XOR (ou exclusif logique) entre deux variables, parmi d'autres, ont pu démontrer que le perceptron à seuil est incapable de résoudre les problèmes de classes non linéairement séparables.

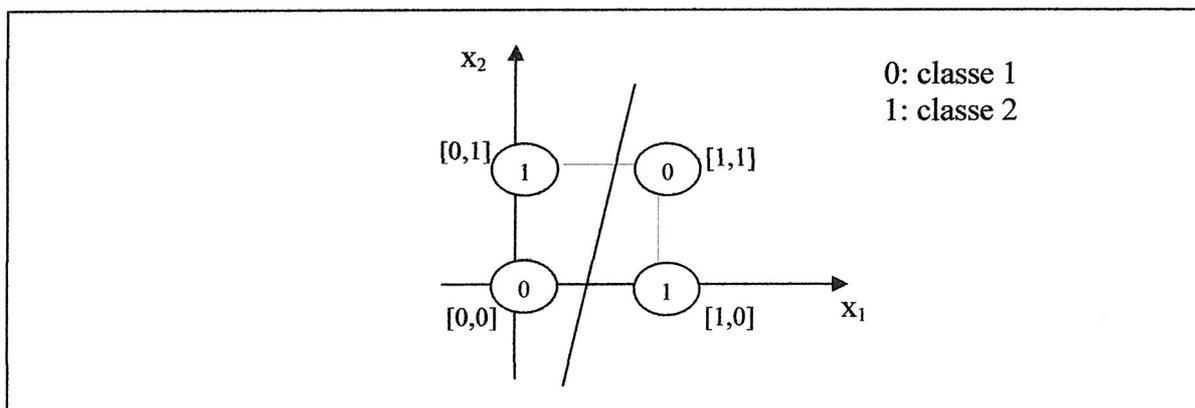


Figure II.7. Exemple du OU exclusif: problème de deux classes non linéairement séparables.

La figure II.7 montre qu'il est impossible de séparer les deux classes à l'aide d'une ligne droite quelque soit son orientation. Aujourd'hui, on sait qu'un perceptron multicouche à une seule couche cachée (figure II.8) parvient à les séparer, seulement à l'époque, on ne disposait pas d'algorithmes d'apprentissage pour le perceptron multicouche, et la démonstration du Minsky et Papert a fortement réduit les recherches sur les réseaux de neurones jusqu'à être délaissés durant près de 20 ans.

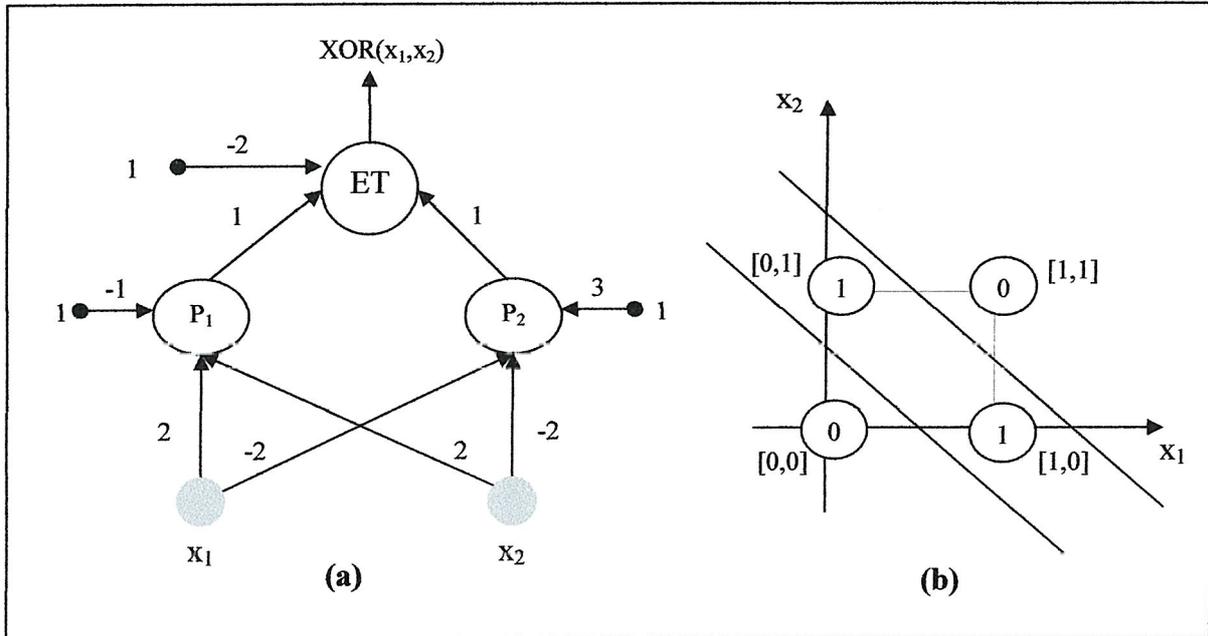


Figure II.8. Implémentation du XOR logique : (a) Calcul du XOR logique par un réseau de neurones à deux couches et (b) Interprétation géométrique de la fonction XOR logique.

I.4 ADALINE et MADALINE (Multiple ADaptive LINEar Element)

Avec une nouvelle règle d'apprentissage, un autre modèle de neurones très important et similaire au perceptron dit ADALINE a été créé par Bernard Widrow et ses étudiants [29]. La différence entre le perceptron et l'ADALINE est que la fonction d'activation Heaviside devient une fonction linéaire et les sorties sont en général réelles. L'ADALINE était appliqué pour le contrôle de systèmes et le traitement adaptatif de signaux (filtre adaptatif pour éliminer les échos des lignes téléphoniques) et aussi pour la prévision des séries temporelles [24].

En remplaçant les perceptrons, dans un réseau de perceptrons, par des neurones ADALINE, on obtient un réseau dit MADALINE présentant donc le même fonctionnement et la même architecture qu'un réseau de perceptrons. Chaque ADALINE partitionne l'espace caractéristique en deux sous-espaces délimités par un hyperplan dont l'équation est la suivante:

$$w_{0m} + \sum_{n=1}^N w_{nm} x_n = 0 \quad \text{avec } m = 1, \dots, M \tag{II.6}$$

et chaque point $X = [x_1 x_2 \dots x_N]$ de cet espace appartenant à l'un des sous-espaces se trouve à une distance ζ_m de cet hyperplan donnée par:

$$\zeta_m = \frac{\left| w_{0m} + \sum_{n=1}^N w_{nm} x_n \right|}{\sqrt{\sum_{n=1}^N w_{nm}^2}} = \frac{|y_m|}{\sqrt{\sum_{n=1}^N w_{nm}^2}} \quad (\text{II.7})$$

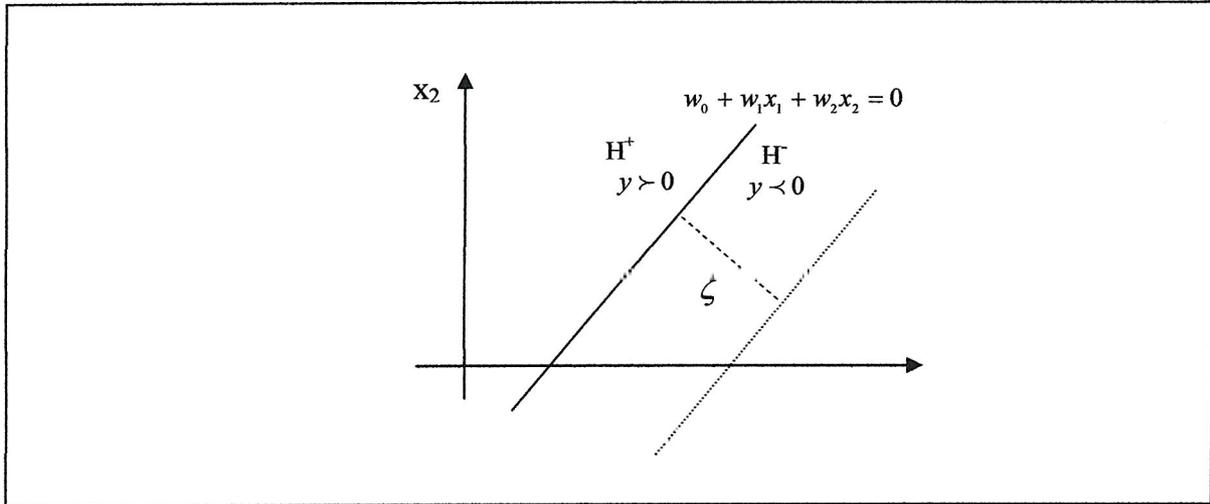


Figure II.9. Classification linéaire par un neurone ADALINE dans le cas bidimensionnel.

Les points ayant la même sortie y_m créent un autre hyperplan parallèle à l'hyperplan défini par l'équation (II.6) et distant de lui de ζ_m . Ce nouvel hyperplan se trouve dans le demi-espace correspondant au signe de la sortie y_m .

Proposée par Widrow et Hoff, la règle d'apprentissage du MADALINE, connue aussi avec le nom LMS (Least Mean Square), consiste à ajuster itérativement les poids du réseau après la présentation de chaque exemple en minimisant l'erreur suivante [15]:

$$E_m(w_{nm}) = \lim_{Q \rightarrow \infty} \frac{1}{Q} \left[\frac{1}{2} \sum_{q=1}^Q (y_m^{(q)} - d_m^{(q)})^2 \right] \quad (\text{II.8})$$

La minimisation de cette erreur peut s'effectuer par la méthode du gradient [15]. Dans ce cas, sa dérivée partielle aura l'expression:

$$\frac{\partial E_m}{\partial w_{nm}} = \frac{1}{Q} \left[\sum_{q=1}^Q x_n^{(q)} \cdot (y_m^{(q)} - d_m^{(q)}) \right] \quad (\text{II.9})$$

et l'adaptation de chaque poids à l'itération $(i+1)$ suit la règle:

$$w_{nm}^{(i+1)} = w_{nm}^{(i)} + \eta \cdot \frac{1}{Q} \left[\sum_{q=1}^Q x_n^{(q)} \cdot (y_m^{(q)} - d_m^{(q)}) \right] \quad (\text{II.10})$$

Widrow et Hoff ont démontré, mathématiquement, que leur algorithme d'apprentissage converge, pour des poids initiaux arbitraires, vers les poids optimaux qui minimisent l'erreur.

I.5 Réseaux de neurones artificiels (Artificial Neural Networks : ANNs)

I.5.1 Définition

Un réseau de neurones artificiel (RNA) est un assemblage de plusieurs neurones formels interconnectés à l'instar des cellules du système nerveux. Les connexions entre les neurones rendent le fonctionnement du réseau adaptatif et parallèle.

Pour le schématiser, un RNA est représenté par un graphe orienté et pondéré formé de nœuds et d'arêtes. Les nœuds représentent les neurones, les arêtes les connexions et les constantes de pondération les poids synaptiques.

I.5.2 Architecture

Les réseaux de neurones artificiels diffèrent selon le nombre de neurones, le type de connexions entre les couches de neurones et la fonction d'activation, ces caractéristiques se regroupent sous le terme « architecture » [24], malgré que nous avons trouvé une confusion entre le terme « architecture » et « topologie ». Il y a ceux qu'utilisent indifféremment architecture ou topologie [15] pour désigner le nombre de neurones et le type de connexions et d'autre qui utilise « architecture » pour le type de connexions seulement. La littérature reporte une grande variété d'architectures de réseaux de neurones artificiels qui dépassent 190 architectures [24], chacune d'elles possède des particularités propres et destinée à des applications spécifiques, mais on peut regrouper les RNA, selon leur type de connexions, en deux grandes classes d'architectures [15]: les réseaux bouclés et les réseaux non bouclés.

I.5.2.1 Réseaux de neurones bouclés (Back-forward , Reccurent Neural Networks)

Dit aussi réseau à architecture cyclique (cyclic), ce type comporte un groupe de neurones interconnectés en anneau: la sortie du 1^{er} neurone est l'entrée du second neurone et la sortie de ce dernier est l'entrée du 3^{ème} ,..., etc. jusqu'au neurone du départ de ce groupe. Si la sortie d'un neurone est l'entrée de tous les autres neurones, on parle de topologie complète [15]. En raison des retards introduits par les boucles et la dépendance de neurones des états précédents, un réseau récurrent est considéré comme système dynamique.

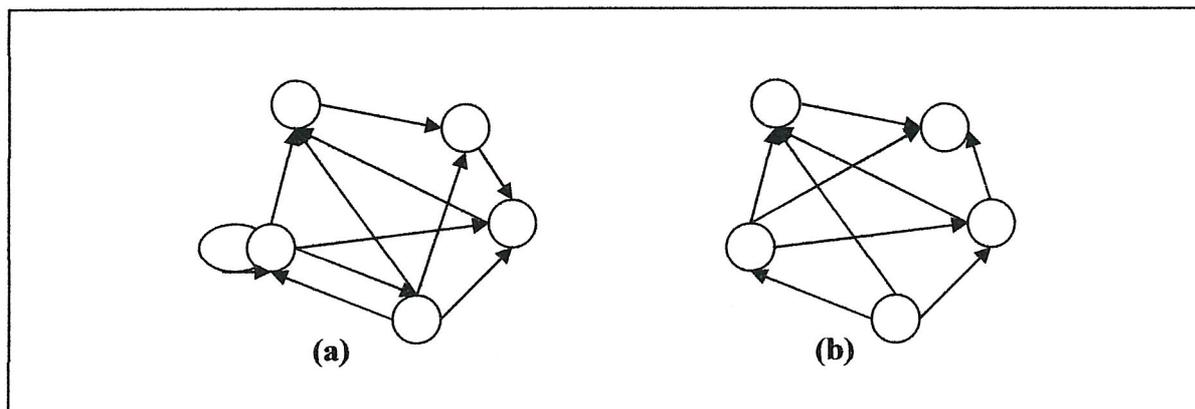


Figure II.10. Architecture des réseaux de neurones: (a) exemple d'un réseau bouclé (architecture cyclique) et (b) exemple d'un réseau non bouclé (architecture acyclique).

I.5.2.2 Réseaux de neurones non bouclés (Feed-forward Neural Networks)

Contrairement à la 1^{ère} architecture, ces réseaux ne présentent aucune boucle (architecture acyclique); il s'agit en fait d'un cas particulier des réseaux bouclés où les poids

récurrents sont nuls [24]. L'information dans un réseau non bouclé suit un sens unidirectionnel: des neurones d'entrée vers les neurones de sortie, c'est pourquoi les neurones peuvent être disposés en couches successives et le réseau est dit alors réseau multicouche comme celui de la figure II.11.

I.5.2.3 Réseaux de neurones multicouche (Multilayered Neural Networks)

Ce type est le plus étudié et utilisé, notamment pour la classification, l'approximation des fonctions et la modélisation des procédés. L'architecture d'un réseau de neurones multicouche (figure II.11) présente des neurones interconnectés et disposés en couches adjacentes telles que les sorties des neurones de la 1^{ère} couche sont les entrées des neurones de la couche adjacente suivante jusqu'à la dernière couche, donc l'information se propage de couche en couche sans retour en arrière. On peut distinguer entre trois types de couches à savoir:

- **Couche d'entrée (Input layer):** c'est la 1^{ère} couche du réseau, elle ne comporte pas de neurones mais sert à transmettre les stimuli externes au réseau,
- **Couche de sortie (Output layer):** est la dernière couche du réseau dont les neurones fournissent les sorties (réponses) du réseau,
- **Couches cachées (Hidden layer) (une ou plusieurs):** sont des couches intermédiaires reliant la couche d'entrée et la couche de sortie et ne présentent aucun contact avec l'extérieur.

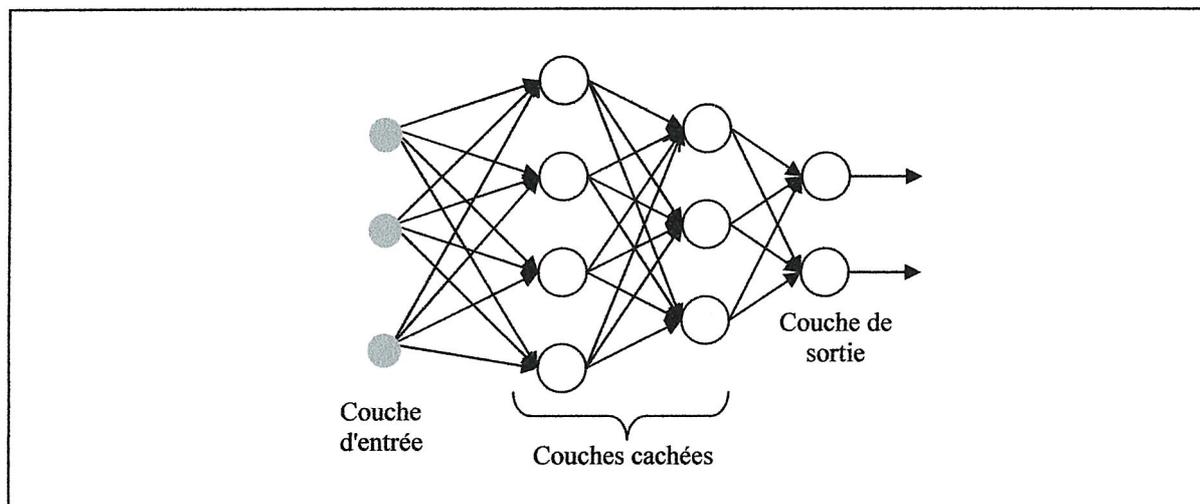


Figure II.11. Réseau de neurones multicouche.

Un réseau de neurones est dit à une seule couche s'il ne possède aucune couche cachée, signifiant qu'il contient une seule couche de poids synaptiques, et il est dit multicouche s'il comporte une ou plusieurs couches cachées.

Par convention [15], on numérote les couches d'un réseau de neurones à partir de 0, qui correspond à la couche d'entrée, jusqu'à la n^{ème} couche (ex. un réseau multicouche à 3 couches comporte une couche d'entrée, 2 couches intermédiaires et une couche de sortie). L'architecture d'un réseau multicouche signifie le nombre de neurones de chaque couche en allant de la 1^{ère} jusqu'à la dernière (ex. l'architecture du réseau de la figure II.11 est 3-4-3-2).

Les connexions, entre les neurones des différentes couches, peuvent être totales ou partielles : pour le 1^{er} cas, la sortie d'un neurone est l'entrée de tous les neurones de la couche suivante, quant aux connexions partielles, elles sont un cas particulier du 1^{er} cas (les connexions absentes possèdent des poids nuls). Au sein d'une même couche, il n'y a pas d'interconnexions.

I.5.2.4 Séparation linéaire par un réseau de neurones multicouche

On va donner une interprétation géométrique de la classification par un réseau multicouche à travers le réseau de la figure II.11. Les neurones de la 1^{ère} couche cachée partitionnent l'espace d'entrée en sous-espaces à l'aide de 4 hyperplans (voir figure II.12) et les neurones de la 2^{ème} couche cachée réalisent un ET logique et fournissent ainsi des régions convexes issues de l'intersection des sous-espaces précédents telles que $K_1 = P_1 \cap P_2 \cap P_3$, $K_2 = P_2 \cap P_3 \cap P_4$ et $K_3 = P_1 \cap P_4$ tandis que les neurones de la couche de sortie engendrent des régions concaves en réalisant un Ou logique autrement dit, chaque neurone de sortie est actif si et seulement si l'une de ses entrées appartient au moins à l'une des régions convexes précédentes, ex : $(K_1 \cup K_2)$.

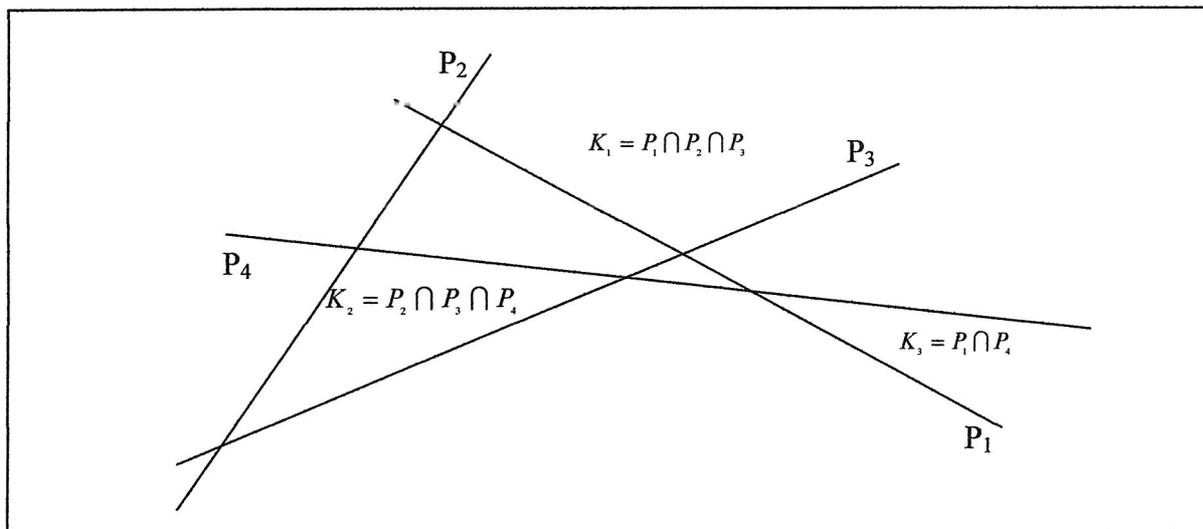


Figure II.12. Exemple de séparation de régions convexes.

I.6 Apprentissage des réseaux de neurones artificiels (Neural Networks Learning)

L'apprentissage d'un réseau de neurones artificiel est une procédure qui consiste à faire un ajustement (adaptation) itératif de ses poids synaptiques en minimisant un certain critère dans le but de réaliser certaines applications telles que la classification.

La procédure d'adaptation est décrite par un algorithme d'apprentissage; il en existe au moins 23 types de règles d'apprentissage [24] dont la 1^{ère} était celle de Hebb dérivée des études psychologiques [31]. Toutes ces règles peuvent être regroupées en trois catégories à savoir: les règles d'apprentissage supervisé, non supervisé et renforcé et suivent les étapes suivantes:

- Présentation d'exemples d'apprentissage,
- Adaptation du comportement du réseau selon ses entrées.

Il est à noter qu'on doit distinguer entre learning (apprentissage) et training (entraînement), ce dernier désigne la présentation des exemples au réseau (et les sorties désirées dans le cas d'un apprentissage supervisé) alors qu'un réseau apprend en s'entraînant.

I.6.1 Apprentissage supervisé (Supervised Learning)

Les réseaux de neurones multicouches adoptent ce mode d'apprentissage qui consiste, après la présentation d'exemples d'apprentissage associés à leur sorties désirées, à minimiser l'erreur entre les sorties du réseau et celles désirées. L'adaptation des poids s'effectue en fonction de cette erreur. Ce mode est utilisé pour la prévision et la classification.

I.6.2 Apprentissage renforcé (Reinforcement Learning)

Une variante de l'apprentissage supervisé est l'apprentissage renforcé pour lequel, on fournit au réseau un score au lieu des sorties désirées, ce score est une mesure du degré de performance du réseau après quelques itérations. Cet algorithme est utilisé, en particulier, dans le domaine des systèmes du contrôle [24].

I.6.3 Apprentissage non supervisé (Unsupervised Learning)

Seul l'ensemble d'exemples est fourni au réseau et les poids sont adaptés de sorte que les exemples similaires sont regroupés dans la même catégorie. L'avantage de ce mode d'apprentissage est son auto organisation [24]. Il est utilisé pour la reconnaissance des patrons, le traitement du signal et l'analyse factorielle.

I.7 Fonctions d'activation (Activation Functions)

La 1^{ère} fonction d'activation des réseaux de neurones artificiels était la fonction Heaviside, proposée pour le neurone formel, mais par la suite de nombreuses fonctions d'activation ont été proposées et qui ne sont pas dérivées des études biologiques mais des mathématiques et d'autres théories telles que la physique [15]. La figure II.13 schématise quelques fonctions d'activations.

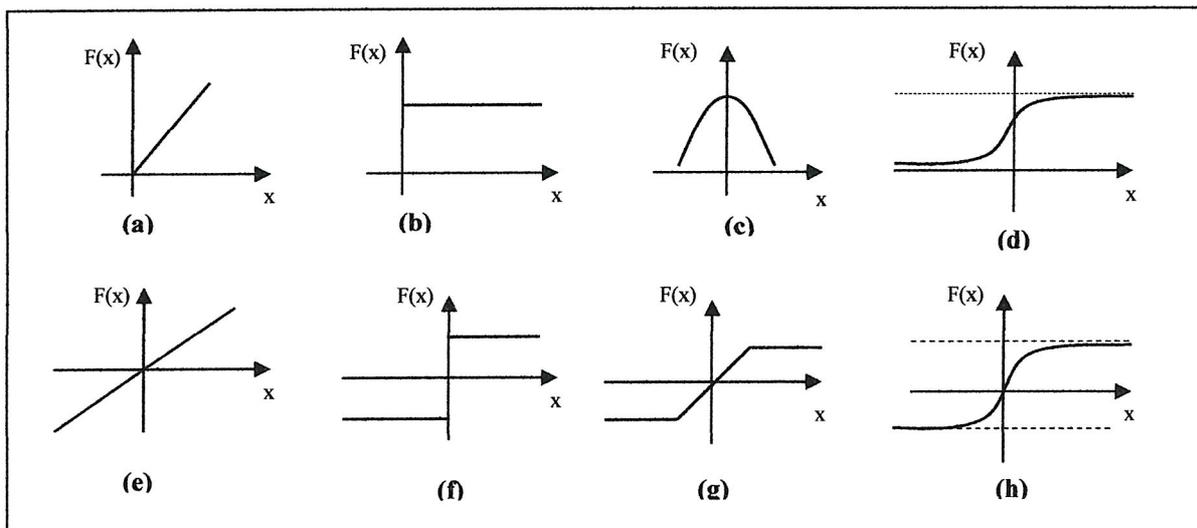


Figure II.13. Différentes fonctions d'activation:

- | | | |
|--------------------------------|-------------------------|------------------------|
| (a) fonction linéaire positive | (d) sigmoïde unipolaire | (g) fonction à seuil |
| (b) fonction Heaviside | (e) fonction linéaire | (h) sigmoïde bipolaire |
| (c) fonction gaussienne | (f) fonction signe | |

II. Le Perceptron Multi-Couche (PMC)

II.1 Présentation du réseau

Le perceptron multicouche (PMC) est l'un des réseaux de neurones les plus populaires, il est une architecture stratifiée de perceptrons; la 1^{ère} couche présente les exemples au réseau, la couche de sortie fournit à l'extérieur les réponses du réseau et une ou plusieurs couches intermédiaires, dites cachées, au sens où ses sorties ne sont pas directement observables.

La figure II.14 schématise un PMC à une seule couche cachée, signifiant qu'il possède deux couches de poids à ajuster. Cette schématisation peut contenir des biais dans chaque couche associés à une entrée additionnelle de valeur «1».

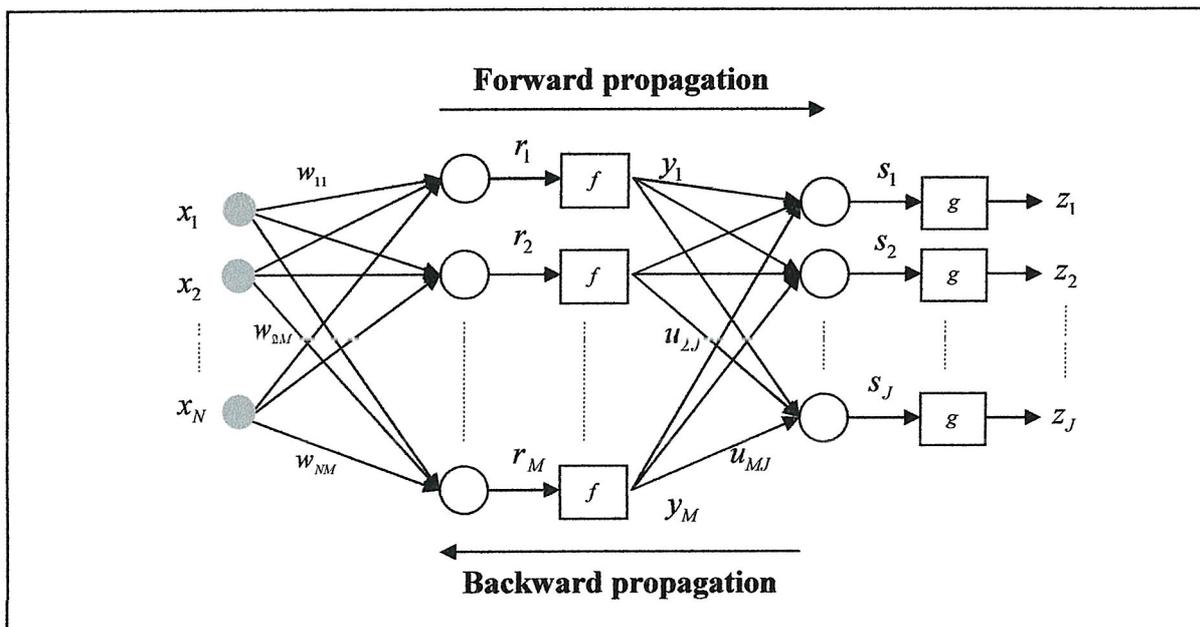


Figure II.14. Schématisation d'un perceptron multicouche.

L'utilisation d'un PMC nécessite qu'on dispose d'un ensemble de Q couples (entrée/sorties désirées), d'un nombre de neurones au sein de la couche d'entrée égal au nombre des attributs (dimension de l'espace caractéristique) et autant de neurones que de classes dans la couche de sortie.

Pour le classifier, un exemple $X^{(q)}(x_1^{(q)} x_2^{(q)} \dots x_N^{(q)})$, présenté à l'entrée du PMC donne, respectivement, dans la couche cachée et celle de sortie les réponses suivantes:

$$y_m = f(r_m) = f\left(\sum_{n=1}^N w_{nm} x_n + w_{0m}\right) \quad \text{avec } m = 1, \dots, M \quad (\text{II.11.a})$$

$$z_j = g(s_j) = g\left(\sum_{m=1}^M u_{mj} y_m + u_{0j}\right) \quad \text{avec } j = 1, \dots, J \quad (\text{II.11.b})$$

s'il appartient à la classe j , le neurone de la sortie correspondante, aura la valeur «1» et les autres auront la valeur «0» ou «-1». Son erreur quadratique, une fonction des poids, est définie par:

$$E^{(q)} = \sum_{j=1}^J (d_j^{(q)} - z_j^{(q)})^2 \quad (\text{II.12})$$

et l'erreur totale de tous les exemples est alors la somme des erreurs quadratiques partielles vérifiant:

$$E_T = \sum_{q=1}^Q \sum_{j=1}^J (d_j^{(q)} - z_j^{(q)})^2 \quad (\text{II.13})$$

II.2 Apprentissage du PMC et algorithme de rétropropagation de l'erreur

L'apprentissage du PMC s'effectue en mode supervisé et vise l'adaptation des poids du réseau en minimisant l'erreur quadratique totale de l'équation (II.13). A cette fin, on utilise un algorithme très populaire dit «rétropropagation de l'erreur», communément appelé «Backpropagation». Cet algorithme, une généralisation de la règle de Widrow-Hoff, et reposant sur la technique de descente du gradient a été, selon [24][27][28], découvert indépendamment par différents chercheurs; en France: Fogelman-Soulié, Gallinari, Le Cun et aux Etats-Unis: Rumelhart, Hinton et Williams pour être appliqué au PMC, mais aujourd'hui, il est utilisé pour différents types de réseaux de neurones et même pour des modèles hybrides Neuro-Fuzzy [24].

L'application de la rétropropagation s'effectue en deux phases: **i) Forward propagation**, dans laquelle, chaque exemple d'apprentissage présenté au réseau est propagé directement vers la couche de sortie via les fonctions d'activation, cette phase renvoie les sorties du PMC et **ii) Backward propagation** pour la comparaison entre les réponses du réseau et celles désirées, et la propagation en arrière de l'erreur, d'où le nom de l'algorithme, et c'est cette erreur qui contribue dans l'adaptation des poids.

Selon l'ajustement des poids, l'algorithme de rétropropagation, se présente en deux versions:

- **Rétropropagation stochastique** ou continue (Stochastic Backpropagation, On-line Backpropagation) : ajuste les poids du PMC après la présentation de chaque exemple d'apprentissage. A l'itération $(i+1)$, l'adaptation des poids de la couche de sortie et de la couche cachée suit, respectivement, les formules suivantes (détaillées dans l'annexe B):

$$u_{mj}^{(i+1)} = u_{mj}^{(i)} + \Delta u_{mj}^{(i)} = u_{mj}^{(i)} + \eta (d_j^{(q)} - z_j^{(q)}) g'(s_j) y_m^{(q)} \quad (\text{II.14.a})$$

$$w_{nm}^{(i+1)} = w_{nm}^{(i)} + \Delta w_{nm}^{(i)} = w_{nm}^{(i)} + \eta \left(\sum_{j=1}^J (d_j^{(q)} - z_j^{(q)}) g'(s_j) u_{mj} \right) (f'(r)_m \cdot x_n^{(q)}) \quad (\text{II.14.b})$$

η , g' et f' désignent respectivement le taux d'apprentissage et les dérivées des sigmoïdes de la couche de sortie et de la couche cachée (équations II.15.a et II.15.b)(voir l'annexe B):

$$g'(s_j) = \alpha \cdot z_j (1 - z_j) \quad (\text{II.15.a})$$

$$f'(r_m) = \alpha \cdot y_m (1 - y_m) \quad (\text{II.15.b})$$

- **Rétropropagation périodique** (Off-line Backpropagation, Batch Backpropagation) : ajuste les poids du PMC une seule fois par itérations après la présentation de tous les exemples, dans ce cas l'accumulation des modifications de chaque exemple contribue à l'ajustement de chaque poids, qui se fait à l'itération $(i+1)$ comme suit:

$$u_{mj}^{(i+1)} = u_{mj}^{(i)} + \eta \sum_{q=1}^Q (d_j^q - z_j^q) \cdot g'(s_j) \cdot y_m^q \quad (\text{II.16.a})$$

$$w_{nm}^{(i+1)} = w_{nm}^{(i)} + \eta \sum_{q=1}^Q \left\{ \left[\sum_{j=1}^J (d_j^q - z_j^q) g'(s_j) u_{mj} \right] \cdot (f'(r_m) \cdot x_n^q) \right\} \quad (\text{II.16.b})$$

Algorithme de rétropropagation de l'erreur

L'algorithme suivant détaille le 1^{er} mode, qui sera utilisé pour nos tests :

Algorithme II.1: On-line Backpropagation

Entrées:

- Nombre d'exemples de la base de données à classifier: Q ,
- Nombre de neurones de la couche d'entrée (dimension du vecteur caractéristique): N ,
- Nombre de sorties du réseau (nombre de classes): J ,
- Vecteurs de sorties désirées pour chaque exemple: d ,
- Taux d'apprentissage: η ,
- Critère d'arrêt.

Sorties:

- Erreur de convergence,
- Taux de classification.

- Initialisation aléatoire des poids et des biais de la couche de sortie: U et b^O
- Initialisation aléatoire des poids et des biais de la couche cachée: W et b^H

Pour $q = 1$ jusqu'à Q

- (A) {
1. Calcul des sorties de la couche cachée et celles de la couche de sortie:
 $Y = f(W^T \cdot X_q^T + b^H)$ et $Z = g(Y^T \cdot U + b^O)$
 2. Calcul de l'erreur quadratique: $E_q = d - Z$
 3. Adaptation des poids et des biais de la couche de sortie:
 $U = U + \eta \cdot Y \cdot (E_q \cdot Z \cdot (1 - Z))$
 $b^O = b^O + \eta \cdot (E_q \cdot Z \cdot (1 - Z))$
 4. Adaptation des poids et des biais de la couche cachée:
 $W = W + \eta \cdot X_q^T \cdot (E_q \cdot Z \cdot (1 - Z) \cdot U^T) \cdot Y \cdot (1 - Y)$
 $b^H = b^H + \eta \cdot (E_q \cdot Z \cdot (1 - Z) \cdot U^T) \cdot Y \cdot (1 - Y)$

Fin pour

5. Calcul de l'erreur quadratique moyenne: $EQM = \frac{1}{Q} \sum_{q=1}^Q \sum_{j=1}^J (d_j^{(q)} - z_j^{(q)})^2$

6. Si le critère d'arrêt est vérifié aller à 7,
 Si non répéter (A)

7. Fin

Remarque : .* désigne le produit terme à terme.

II.3 Problèmes de l'algorithme de rétropropagation de l'erreur

Malgré, la popularité et la puissance du Backpropagation, notamment pour le PMC, qui est utilisé avec succès dans près de 80% des applications des réseaux de neurones artificiels [15], cet algorithme est consommateur de temps de calcul sur des problèmes de tailles et souffre: **i)** de la convergence, à partir des poids initiaux et un taux d'apprentissage bien choisi, vers le minimum local de la fonction erreur, **ii)** de la lenteur de cette convergence et **iii)** de l'absence d'un critère d'arrêt unique.

II.3.1 Minima locaux

Les algorithmes de type descente du gradient souffrent de la convergence vers le minimum local de la fonction à minimiser et c'est uniquement par chance qu'ils convergent vers un minimum global. Pratiquement, pour remédier à ce problème, on applique l'une de ces solutions [14][20]: **i)** réalisation de plusieurs apprentissages avec différentes initialisations aléatoires ; cette procédure ne garantit pas la convergence vers le minimum global, mais augmente la probabilité d'avoir une solution suffisamment bonne, cependant, on ne peut pas l'appliquer dans le cas des problèmes de taille où l'apprentissage peut durer des jours ou des semaines, **ii)** utilisation de la technique «shaking the weights»; c'est-à-dire ajouter du bruit aux données ou **iii)** l'augmentation, dans le cas de fonctions ayant plusieurs minima locaux, de la taille du réseau en rajoutant des neurones cachés et/ou de couches cachées. En résumé, on se contente d'aboutir, non pas à un minimum global, mais à un minimum local acceptable.

II.3.2 Lenteur de la convergence

Nombreuses, sont les méthodes qui visent à accélérer la phase d'apprentissage du PMC. Parmi lesquelles, on compte l'ajout d'un terme d'inertie dit moment (momentum) [34], pour lequel, l'adaptation des poids, suit la règle suivante:

$$w_{nm}^{(i+1)} = w_{nm}^{(i)} + \eta \cdot \frac{\partial E^{(i)}}{\partial w_{nm}} + \mu(w_{nm}^{(i)} - w_{nm}^{(i-1)}) \quad (\text{II.17})$$

Où μ est une petite constante qui désigne le moment ($0 < \mu < 1$).

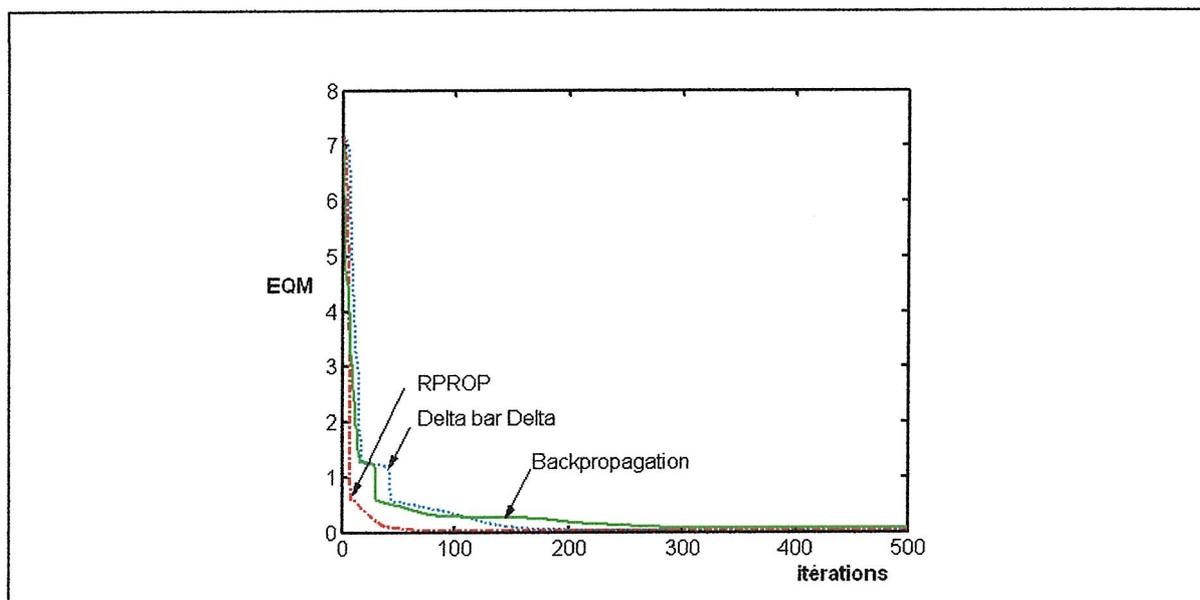


Figure II.15. Comparaison entre différentes techniques d'accélération de l'apprentissage.

et l'utilisation du mode batch pour l'adaptation des poids en intégrant des algorithmes d'accélération de la convergence tels que Quickprop [13], RPROP [21] et la règle delta bar delta [33]. Ces algorithmes assignent à chaque poids son propre taux d'apprentissage qui s'ajuste itérativement lors de l'apprentissage.

Nous avons comparé trois algorithmes à savoir : Backpropagation, RPROP et la règle delta bar delta pour illustrer l'effet de l'accélération de l'apprentissage. Les résultats de l'application sur la base 10-5-10 et avec les mêmes poids initiaux sont illustrés en figure II.15.

II.3.3. Critère d'arrêt

Pour arrêter l'apprentissage de l'algorithme de rétropropagation de l'erreur, plusieurs critères sont utilisés et qui dépendent de l'application que l'on traite [14]:

- Fixation d'un nombre maximal d'itérations; ce critère assure l'arrêt de l'algorithme même s'il est pris dans un minimum local ou si les paramètres du réseau ne sont pas bien choisis,
- Fixation d'une borne inférieure sur l'erreur quadratique moyenne (MSE) ou sur sa racine,
- Arrêt si la variation de l'erreur ou des poids devient très lente,
- Fixation d'une limite inférieure sur la composante absolue maximale $e_i^{(q)}$ du vecteur de l'erreur $[e_1^{(q)} e_2^{(q)} \dots e_N^{(q)}]$,
- Utilisation d'un ensemble de validation (ensemble indépendant de l'ensemble d'apprentissage et de test): l'algorithme est arrêté lorsque les performances du réseau, testé sur cet ensemble, cessent de s'améliorer. Dans ce cas, on préserve les poids optimaux obtenus en utilisant l'ensemble de validation et non pas les poids obtenus avec l'ensemble d'apprentissage.

II.4 Saturation des neurones

Un neurone est dit saturé si les exemples à classer se situent dans les régions non linéaires de la sigmoïde correspondant aux valeurs de sa dérivée qui tendent vers 0 (voir figure II.16). Pour ces valeurs, l'ajustement des poids sera négligeable et l'apprentissage très lent [13][14], autrement dit les neurones cessent d'apprendre. D'autre part, les poids initiaux ayant de grandes valeurs provoquent aussi la saturation des neurones [14][16][21].

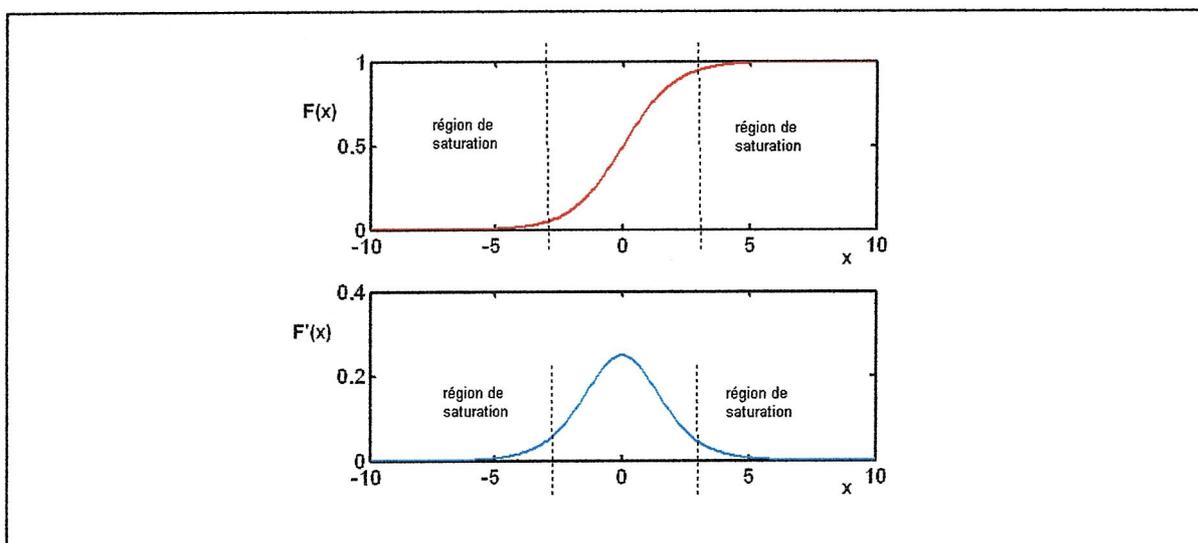


Figure II.16. Illustration des régions de saturation des neurones.

Parmi les solutions proposées pour remédier à ce problème, celle de Fahlman [13] qui consiste à ajouter une petite constante de valeur 0.1 à la dérivée de la sigmoïde, et celle de la modification des sorties désirées du réseau [14][16] avec de nouvelles valeurs [0, 1.1],[0.05, 0.95] pour la sigmoïde ou [-1.1, 1.1],[-0.9, 0.9] pour la tangente hyperbolique. Pour le cas d'un important fan-in (nombre des entrées du neurone : d_{in}), on choisit d'initialiser les poids dans l'intervalle [-a,+a] tel que le rang « a » soit inversement proportionnel au fan-in [13] ou vérifie [14]:

$$a = k/d_{in} \quad \text{avec } k \geq 1 \quad \text{ou } k \in [0.01, 0.1]$$

II.5 Généralisation

Un réseau de neurones généralise bien s'il arrive à performer avec des données qu'il n'a pas rencontrés durant la phase d'apprentissage, dans le cas contraire, on dit que ce réseau ne possède pas la capacité de généralisation. Cette capacité a été remarquablement illustrée par le réseau «NET-Talk» qui a pu prononcer de nouveaux mots après un apprentissage de 12 heures [28]. En effet, le but de l'apprentissage est de réaliser un réseau doté du pouvoir de généralisation et non pas un réseau performant sur l'ensemble d'apprentissage seulement [14].

La généralisation dépend des poids initiaux [17], de l'architecture du réseau [19] et de la taille de l'ensemble de données [14]. Ce dernier doit être représentatif du problème traité, et de taille appropriée car on ne peut pas entraîner un grand réseau avec quelques exemples seulement.

Eviter le problème de saturation des neurones lors de l'apprentissage et réaliser un PMC doté du pouvoir de généralisation, exigent une bonne connaissance du choix de l'architecture du réseau et des paramètres intervenants dans l'algorithme de Backpropagation. A cet objectif, nous réservons les paragraphes suivants.

II.6 Choix des paramètres du PMC

Les poids initiaux, le taux d'apprentissage et sa technique d'adaptation, le gain de la fonction d'activation influencent énormément sur la vitesse de la convergence, le temps d'apprentissage et la généralisation; c'est pourquoi plusieurs méthodes et techniques ont été proposées ; soit pour l'initialisation des poids, soit pour l'adaptation du taux d'apprentissage, et d'autres études comparatives pour l'évaluation des méthodes d'initialisation [16][17] et l'adaptation du taux d'apprentissage [18][19] ont été réalisées.

Il n'existe pas de méthodes automatiques pour déterminer les paramètres optimaux du réseau réalisant, par exemple, un temps d'apprentissage court et une bonne généralisation. En pratique, ces paramètres sont déterminés après avoir effectué plusieurs expériences. Durant la recherche d'une bonne combinaison de ces 3 paramètres, il suffit [16] de fixer un paramètre et de varier les deux autres.

II.6.1 Fonction d'activation

La littérature reporte de nombreuses fonctions d'activation, mais le PMC, dédié à la classification, exige des fonctions d'activation : **i)** non linéaires pour pouvoir générer des frontières de discrimination non linéaires et **ii)** continues et dérivables pour que l'algorithme de rétropropagation puisse fonctionner. Pratiquement, les fonctions d'activation des couches du PMC peuvent être différentes mais au sein d'une même couche, on utilise des fonctions d'activation identiques [27].

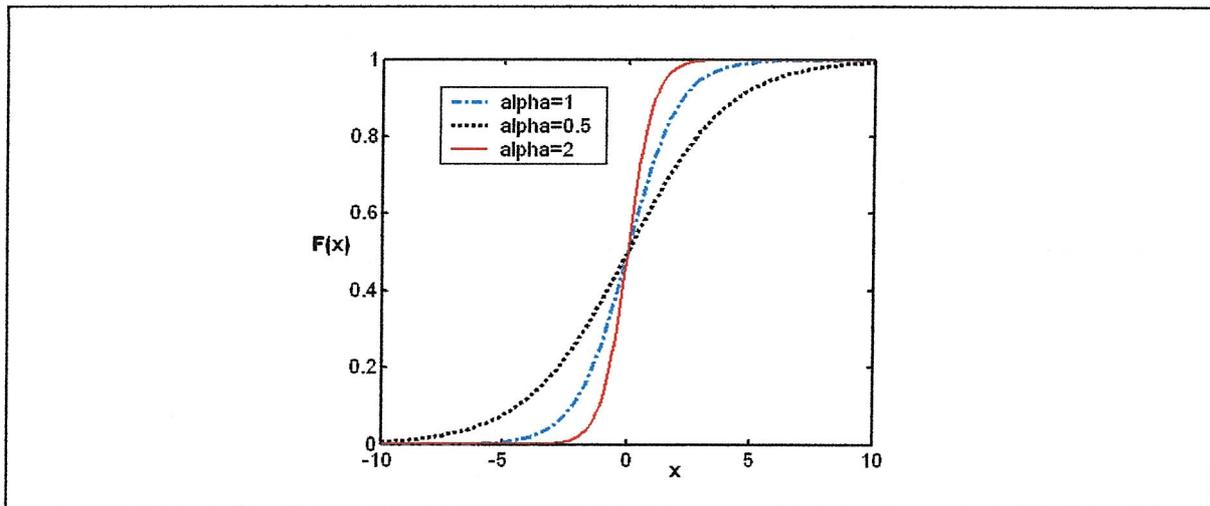


Figure II.17. Fonctions sigmoïdes en fonction du gain.

Le choix usuel est la famille des sigmoïdes présentant la forme d'un «S» et ayant l'expression générale suivante:

$$F(x) = \frac{\lambda}{1 + e^{-\alpha x} + \beta} + \delta \quad (\text{II.18})$$

Pour $\lambda = 2, \alpha = 1, \beta = 0$ et $\delta = -1$, on obtient une sigmoïde bipolaire,

Pour $\lambda = 1, \alpha = 1, \beta = 0$ et $\delta = 0$, on obtient une sigmoïde unipolaire (voir figure II.17).

Le paramètre « α », sur lequel dépend la convergence de l'algorithme de rétropropagation, est dit gain de la sigmoïde. En général, il est mis à 1, mais il peut être spécifié pour chaque neurone et la sigmoïde devient alors:

$$F_j(x) = \frac{1}{1 + e^{-\alpha_j x}} \quad \text{avec } j \text{ l'indice du } j^{\text{ème}} \text{ neurone} \quad (\text{II.19})$$

II.6.2 Initialisation des poids (Weight initialisation)

Pour initialiser les poids du PMC, plusieurs méthodes ont été proposées et que l'on peut regrouper en deux catégories [16]: **i)** initialisation aléatoire et **ii)** initialisation avec des valeurs bien déterminées. Cette dernière est coûteuse en temps de calcul et nécessite une analyse statistique et/ou géométrique de données, parmi lesquelles, la méthode pseudo-inverse. L'initialisation aléatoire a été proposée par Rumelhart et al. (1986) après avoir remarqué que des poids initiaux nuls donnent des poids identiques durant la phase d'apprentissage. Par la suite plusieurs chercheurs l'ont adoptée, c'est pourquoi nous allons réserver un espace de ce mémoire pour donner un bref aperçu, résumé à partir de [16], de ces techniques.

Fahlman [13] a utilisé, pour ses tests, des poids aléatoires qui suivent une loi uniforme dans l'intervalle $[-a, +a]$. Le rang « a » était déterminé empiriquement, cependant d'autres chercheurs l'ont fixé en fonction du fan-in (nombre des entrées du neurone):

L.Bottou a pris $a = -r/\sqrt{d_{in}}$ où $r = 2.38$ pour la sigmoïde et 0.66 pour la tangente hyperbolique et J.Bores et H.Kuiper (1992) ont choisi $-3/\sqrt{d_{in}}$ tandis que F.J.Smieja a fixé, pour chaque neurone, la valeur $2/d_{in}$. Y.Kim et J.Ra (1991) ont fixé une limite inférieure des poids à η/d_{in} où η est le pas d'apprentissage.

On note également, la méthode mixte de C.Chen et R.Nutter (1991), pour le PMC, combinant une initialisation aléatoire pour les poids de la couche cachée et la méthode pseudo-inverse pour les poids de la couche d'entrée et celle de Drago et Riedella (1992) [17] où les poids de la couche d'entrée et ceux de la couche cachée sont initialisés respectivement selon :

$$w_{nm} = \frac{1.3}{\sqrt{1 + N_{input} \cdot v^2}} r_{nm} \quad \text{et} \quad w_{nm} = \frac{1.3}{\sqrt{1 + N_{hidden} \cdot v^2}} r_{nm} \quad (\text{II.20})$$

où r_{nm} une variable aléatoire uniformément distribuée dans l'intervalle $[-1,1]$ et v la moyenne carrée des entrées.

Le choix des poids initiaux est critique [13][16][19] compte tenu que de petites valeurs ralentissent la convergence et nécessitent beaucoup d'itérations pour atteindre l'erreur minimale et dans les cas pires, l'erreur cesse de diminuer et l'algorithme reste piégé dans un minimum local indésirable. D'autre part, les grandes valeurs accélèrent la convergence mais provoquent la saturation des neurones du PMC.

Pour illustrer l'effet des poids initiaux sur la convergence de l'erreur, nous avons entraîné un PMC avec l'algorithme de Backpropagation et initialisé aléatoirement les poids dans deux intervalles $[-0.25, 0.25]$ et $[-0.5, 0.5]$. Les figures II.18(a) et II.18(b) représentent, respectivement, l'évolution de l'erreur quadratique moyenne obtenue pour les deux bases : 10-5-10 et Iris de Fisher.

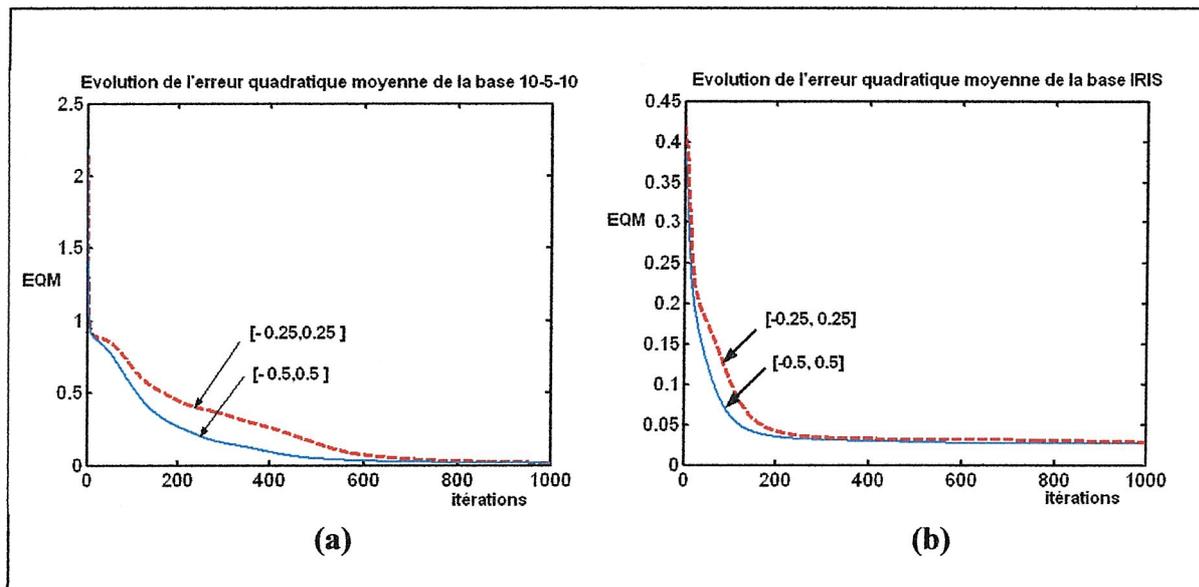


Figure II.18. L'effet des poids initiaux sur la convergence pour :
(a) la base 10-5-10 et (b) la base IRIS.

II.6.3 Taux d'apprentissage (Step Size learning, Rate learning)

Nombreuses sont les techniques qui déterminent le pas d'apprentissage (dit aussi taux d'apprentissage), mais on peut les classer en deux catégories [22]: **i) les techniques d'adaptation globale** pour lesquelles le taux d'apprentissage est unique pour tous les poids et s'adapte à chaque itération. Ces techniques exploitent des connaissances générales sur le réseau entier telles que la direction des poids adaptés et **ii) les techniques d'adaptation locale** qui assignent un taux d'apprentissage individuel à chaque poids. Ces dernières visent à accélérer le mode d'apprentissage batch en trouvant les taux optimaux de chaque poids.

Nous résumons à partir de [18] quelques techniques d'adaptation globale:

- La méthode de Harry (1992) où le taux d'apprentissage η est une constante déterminée pour le mode batch selon:

$$\eta = \frac{1}{\sqrt{N_1^2 + N_2^2 + \dots + N_p^2}} \quad (\text{II.21})$$

après avoir subdiviser l'ensemble de données en P sous ensembles de tailles N_p avec $p = 1, \dots, P$

- Search-Then-Converge proposée en 1990 par C.Darken et J.Moody qui consiste, pour le mode on-line, à diminuer le taux d'apprentissage η à l'itération (i) somme suit:

$$\eta^{(i)} = \frac{\eta^{(0)}}{1 + \frac{i}{r}} \quad (\text{II.22})$$

Où $\eta^{(0)}$ désigne le taux d'apprentissage initial avec une grande valeur, et r est une constante déterminée à partir de l'erreur et des itérations.

On cite également, la méthode «Angle Driven Learning» proposée par L.Chan et F.Fallside (1987) qui détermine le taux d'apprentissage η à partir de l'angle mesuré entre $\Delta E(i)$ et $\Delta w(i-1)$, et la combinaison du gradient conjugué avec «Search Line» proposée par J.Leonard et M.Kramer en 1990.

Quant aux méthodes de la deuxième classe, leur principe est d'augmenter le pas d'apprentissage si les changements successifs des poids sont dans la même direction et le diminuer autrement. Elles utilisent des coefficients heuristiques pour assurer la convergence vers le minimum et éviter les oscillations [19]. Parmi ces méthodes, on compte l'algorithme RPROP proposé par Riedmiller et Braum [21], QuickPROP de Fahlman [13], la règle Delta-Bar-Delta proposée par Jacobs [33], Sign Changes de Silva et Almeida et SuperSAB de Tollenaere [22]. Vu l'importance de l'accélération du mode batch, ces méthodes étaient un sujet d'études comparatives dans [18][19][22].

II.6.4 Taille du perceptron multicouche

Il y a une relation entre la taille du réseau (nombre de paramètres, nombre de neurones), la taille de la base de test et l'erreur de validation [20], si le réseau est trop grand, il y aura un risque d'un surapprentissage et donc d'une généralisation médiocre et s'il est trop

petit, il n'arrive pas à performer sur l'ensemble d'apprentissage [18][20]. On peut regrouper les techniques de détermination de l'architecture du réseau multicouche en trois catégories [4][14]:

- **Les techniques de pruning (Pruning techniques)** qui consistent à commencer avec une architecture grande et la diminuer itérativement en éliminant des connexions ou de neurones entier jusqu'à l'obtention d'une architecture adéquate. Parmi ces méthodes, on cite les méthodes « Skeletomization », «Optimal brain damage » et «Optimal brain surgeon ». Malgré l'efficacité de ces techniques, elles sont coûteuses en temps de calcul,
- A l'opposé de la 1^{ère} catégorie, **les techniques constructives (Constructives techniques)** débutent avec un réseau de petite taille et d'autres neurones seront rajoutés progressivement jusqu'à atteindre une performance acceptable. Une technique connue pour son efficacité [18] est Correlation-Cascade proposée par Fahlman et Lebiere en 1990,
- **La méthode analytique** emploie des techniques statistiques et algébriques pour déterminer le nombre M de neurones de la couche cachée. Le PMC à une seule couche cachée partitionne l'espace caractéristique de dimension N en J régions polyédrales telles que :

$$J = \sum_{n=0}^N \binom{M}{n} \quad \text{avec} \quad \binom{M}{n} = 0 \quad \text{pour} \quad M < n \quad \text{ct} \quad \binom{M}{n} = \frac{M!}{n!(M-n)!}$$

II.7 Applications des réseaux de neurones pour la détection et la reconnaissance de visages

L'auto adaptabilité, la généralisation, le parallélisme, la tolérance aux fautes et au bruit, le réapprentissage (re-learning) et l'universalité ont fait des réseaux de neurones des systèmes très attractifs dans divers domaines tels que la reconnaissance de formes (chiffres et caractères manuscrits ou imprimés,...), reconnaissance de la parole, la prédiction non linéaire, la modélisation, la prévision hydraulique,... Parmi les exemples les plus populaires et aussi spectaculaires, on cite le réseau NET-Talk qui apprend à lire à haute voix des textes en anglais, le réseau ALVINN (automobile autopilote) qui consiste en une voiture guidée par un réseau de neurones, le système SNOOP pour la détection d'explosives plastiques dans les bagages des voyageurs et les réseaux de la société NESTOR pour la reconnaissance de codes postaux et l'identification de cibles et de bruit,....

Nous allons se focaliser sur la reconnaissance de visages, notre axe de recherche, et la détection de visages pour lesquels la littérature reporte énormément de méthodes basées sur les approches connexionnistes.

L'un des premier systèmes de reconnaissance de visages utilisant les réseaux de neurones artificiels est le système WISARD (Wilkie, Stonham et Alexandre's Recognition Devise) [105] réalisé en 1986 à des fins de reconnaissance et de vérification. Ce réseau qui consistait en un réseau à une seule couche cachée pour chaque classe arrivait remarquablement à reconnaître l'identité et les expressions faciales de personnes, seulement, il dépendait de l'éclairage, des arrières plans et de la taille des images de visages. Pour une base de 200-400 images de 16 personnes, le taux de reconnaissance du WISARD était 100%.

Un autre système est le système **GANNET** qui utilisait les algorithmes génétiques pour spécifier la structure du réseau de neurones entraîné par l'algorithme de rétropropagation. Il utilisait les caractéristiques géométriques et topographiques pour représenter les visages.

Kohonen en 1989, Kohonen et Lahito en 1981 ont classifié les visages par un réseau de neurones associatif. En se basant sur leur travail, Fleming et Cothell, en 1990, ont utilisé un réseau entraîné par la rétropropagation pour la reconnaissance du genre (gender).

En 1997, Lawrence et al. ont proposé un réseau de neurones hybride qui combine le réseau SOM pour l'extraction de caractéristiques et un réseau convolutionnel pour la classification. Le taux de reconnaissance obtenu pour la base ORL était 96.2%.

En 1997, Lin et al. [86] ont utilisé un système neuronal entièrement automatique basé sur une décision probabiliste appelé «PDBNN». Il peut détecter les visages dans une scène encombrée, localiser les yeux et reconnaître les visages presque en une seconde avec un taux de 96%.

Baron [104] a proposé en 1981 un système semi automatique capable de reconnaître des visages orientés de 20°. Pour ce faire, il a utilisé le visage entier et ses quatre régions (yeux, bouche, menton et cheveux), arrivant ainsi à reconnaître à 100% les 42 personnes formant sa base de visages.

Récemment, pour améliorer la performance de leur système de vérification (authentification) de personnes à base d'un perceptron multicouche (PMC), Marcel et Bengio [75] ont combiné le vecteur-visage et l'information couleur. Quant à T-P.Nguyen et T-L.Nguyen [79] ont fourni au PMC un vecteur formé de 6 distances (entre les yeux, entre les yeux et le nez et entre les yeux et la bouche).

La majorité des systèmes connexionnistes, dédiés à la reconnaissance de visages, considèrent le visage comme une forme 2D et le réseau reçoit ainsi en entrée un vecteur qui représente toute l'image du visage et/ou des régions de ce dernier. Les inconvénients de ces systèmes sont dus à la lenteur de leur phase d'apprentissage et à la sensibilité aux changements d'illumination ainsi qu'à la nécessité d'un nombre important d'exemples pour l'apprentissage afin d'atteindre une performance acceptable.

Pour la détection et la localisation de visages, on trouve dans [73] plusieurs méthodes basées sur les réseaux de neurones artificiels, parmi lesquelles, les réseaux associatifs et les réseaux convolutionnels, mais le système ayant suscité beaucoup d'intérêt est celui de Rowley et al. [36]. Testé sur une base de 130 visages, ce système arrivait à détecter 90.5% de visages, seulement il était limité à la détection des visages de face, c'est pourquoi, les mêmes chercheurs ont proposé « Router Network » pour la détection des visages orientés [37].

Conclusion

Dés la création des réseaux de neurones, en 1943, les chercheurs ne cessent d'apporter des améliorations que ce soit pour l'algorithme d'apprentissage, les techniques d'initialisation des poids, les méthodes de construction des réseaux ou pour l'adaptation du taux d'apprentissage. Malgré tous ces efforts, le PMC entraîné par l'algorithme Backpropagation reste souffrant, entre autres, de la lenteur de sa phase d'apprentissage et de la convergence vers des minima locaux, mais ça n'empêche pas de dire qu'il a eu le temps de faire ses preuves comme classificateur doté d'un pouvoir particulier; la généralisation.

Notre analyse et passage en revue de la littérature, nous les espérons utiles aux lecteurs désirant utiliser un PMC pour la classification.

Chapitre III

Biométrie & Reconnaissance de Visages

Introduction

Présenter son badge avant d'accéder à son lieu de travail, signer avant toute transaction, composer son code confidentiel avant d'ouvrir son coffre fort, taper son mot de passe avant d'utiliser son ordinateur,... sont des actes d'identification. Mais pourquoi et comment s'identifier ? Comment s'assurer que la personne qui déclare être X est réellement celle qui le prétend être ? Les réponses à toutes ces questions, parmi d'autres, seront l'objectif de la première partie de ce chapitre dédiée à la biométrie et l'identification de personnes.

Inspirés par le système visuel humain, des chercheurs ont tenté depuis les années soixante dix de recopier les mécanismes de la perception chez les humains et de réaliser des systèmes capables de reconnaître une personne à partir de son visage. Ils ont réussi, avec une certaine fiabilité, en adoptant plusieurs méthodes évaluées sur différentes bases de visages. Ainsi, la deuxième partie de ce chapitre est dévolue à présenter une analyse et un état de l'art de ces systèmes de reconnaissance faciale. A cette fin, la perception du visage est traitée du point de vue neurophysiologique avant d'être abordée du point de vue scientifique.

I Biométrie et Identification d'Individus

I.1 Identification d'individus

Nous arrivons à bien reconnaître les personnes que nous côtoyons à partir de leur visage, leur voix, leur démarche, leur style vestimentaire, leur style de coiffure, leur silhouette... mais, pour un agent par exemple, identifier une personne au guichet d'une banque, à l'aéroport,... ou qui essaie d'accéder à un laboratoire de recherche ou un immeuble surveillé, nécessite d'autres moyens d'identification efficaces et rapides. A cette fin, d'autres moyens sont utilisés, et tout individu est alors identifié, soit :

- par ce qu'il possède (Token-based) telles qu'une carte d'identité, un badge, une clé,...
- par ce qu'il sait (Knowledge-based): un mot de passe, un numéro d'identification personnelle (NIP),...
- ou par ce qu'il est, il s'agit de caractéristiques biométriques qui sont uniques et propres à lui. Ces caractéristiques ne peuvent être ni volées ni perdues et sont très difficiles à mimer ce qui rend cette méthode une alternative aux deux précédentes.

I.2 Biométrie (Biometric)

I.2.1 Définition : Le terme « biométrie » est à l'origine grecque: « bio » désigne vie et « metron » mesure, il remplace le terme « anthropométrie » venant de « anthropos » qui signifie homme et « metrie » pour mesure désignant ainsi les mesures du corps humain. La biométrie est donc une technologie de vérification d'identité de personnes basée sur la connaissance de:

- **Caractéristiques morphologiques** ou physiologiques d'un individu : visage, empreintes digitales, iris, rétine, géométrie de la main, géométrie de l'oreille,...
- **Caractéristiques comportementales** telles que la signature manuscrite, la démarche, le rythme de frappe sur un clavier d'ordinateur,...
- **Caractéristiques biologiques** telles que l'ADN, l'odeur corporelle, la salive, le sang,...

avec leur diversité, ces caractéristiques biométriques doivent être [40][43] universelles, uniques, permanentes, collectables, mesurables, discriminantes, difficilement falsifiables et acceptables par les utilisateurs.

I.2.2 Domaines d'application de la biométrie

Le besoin de moyens rapides et précis d'identification avec les capacités de calcul offertes par des ordinateurs puissants, le champ des applications de la biométrie s'élargit de jour en jour, pour recouvrir plusieurs domaines. Parmi les applications, on cite :

- **le contrôle d'accès (Acces control):** il peut être physique à une salle informatique ou virtuel à une source ou un service comme l'accès aux réseaux d'ordinateurs et aux sites sensibles,
- **la répression (Law enforcement):** une des applications à la répression est la criminologie,

- **la personnalisation (Personalization)** des appareils que nous utilisons tous les jours comme les voitures [43],
- **l'authentification de transactions (Transactions authentication)**: elle englobe le retrait d'argents aux guichets des banques, les paiements par cartes bancaires, les transferts de fond, et les paiements effectués à distances.

I.3 Systèmes biométriques : principe et modes de fonctionnement

Les systèmes biométriques se basent sur le même principe [39][43], illustré en figure III.1, pour effectuer deux types de comparaison à savoir:

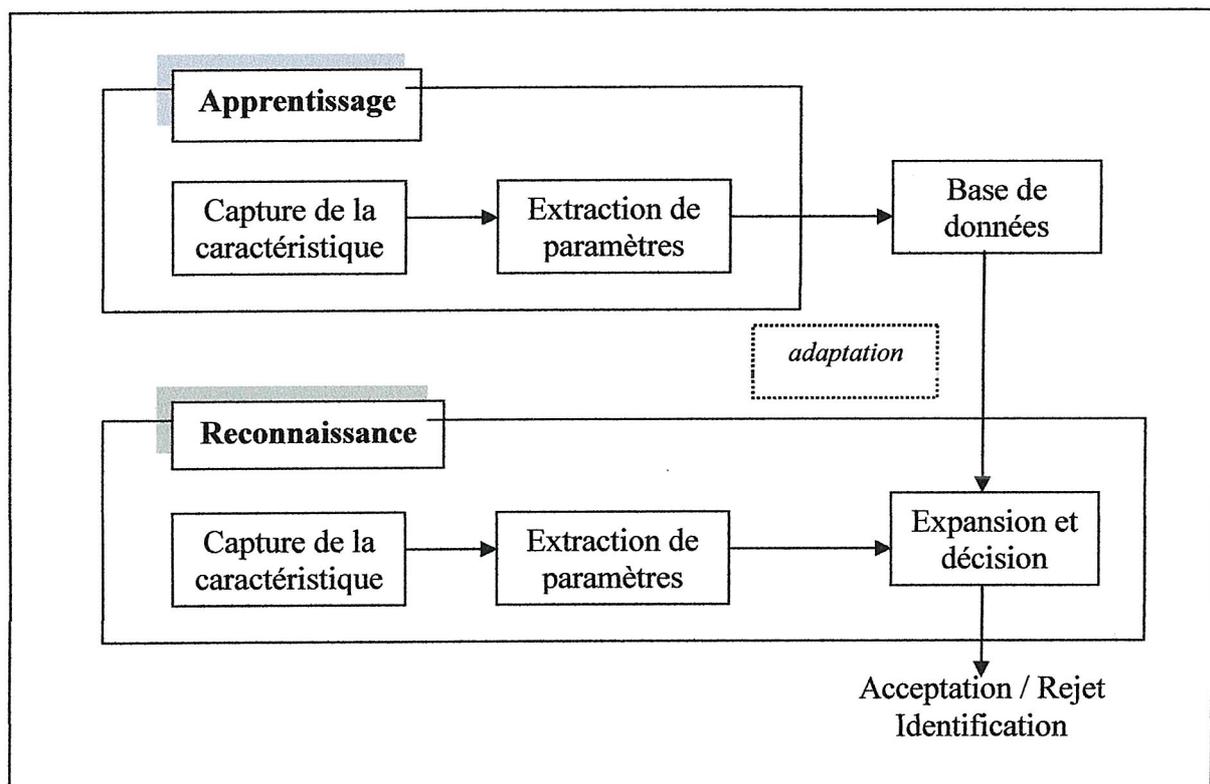


Figure III.1. Architecture d'un système de reconnaissance biométrique.

- **un à plusieurs**: pour répondre à la question «qui suis-je ? »; ces systèmes visent à établir l'identité d'une personne en comparant ses caractéristiques avec celles de la base de données. Ce mode est considéré «fermé» c'est-à-dire que toute personne qui utilise ce système possède un modèle dans la base de données,
- **un à un**: pour effectuer une vérification, appariement ou authentification biométrique et répondre à la question: « suis-je effectivement la personne que je déclare ? » et cela en comparant un gabarit de vérification avec un autre d'enrôlement. Ce mode est «ouvert» car il y a des personnes ne possédant pas de modèles dans la base de données (imposteurs) qui cherchent à être reconnus.

Pour ces deux modes de fonctionnement, les systèmes biométriques peuvent se répartir, selon qu'ils nécessitent ou non la coopération de la personne à identifier, en systèmes

intrusifs (systèmes basés sur l'iris, la rétine, l'ADN,...) et systèmes non intrusifs (basés sur le visage, la démarche,...)

I.4 Classification des systèmes de reconnaissance biométrique

Pour mener à bien une tâche d'identification ou d'authentification, les systèmes biométriques doivent être choisis en fonction de l'application désirée et des données disponibles. Tous les systèmes possèdent des avantages et des inconvénients et le choix de l'un d'eux nécessite des critères d'évaluation et de classification. A cet objectif, on se réfère aux caractéristiques suivantes [40]:

1. **la finalité:** distinction entre l'objectif d'identification et celui de vérification/authentification,
2. **le mode d'activation:** selon que l'activation de la reconnaissance est réalisée par la personne à reconnaître (ou au moins sa présence) ou non,
3. **la tolérance par l'utilisateur:** certaines techniques biométriques sont plus acceptées par les utilisateurs que d'autres telles que la reconnaissance faciale et vocale,
4. **la robustesse:** cette qualité caractérise la résistance à la falsification ou à l'imposture,
5. **l'interfaçabilité :** est la qualité d'un système qui peut être utilisé avec d'autres systèmes informatiques.
6. **le stockage:** la possibilité de coupler la base de données à d'autres informations, de la transmettre à d'autres systèmes pour d'autres utilisations,
7. **la performance :** elle se mesure en termes d'erreur et de vitesse d'exécution. Pour les mesures d'erreur, on doit distinguer entre les systèmes dédiés à la vérification (authentification) et ceux d'identification [39][43]:

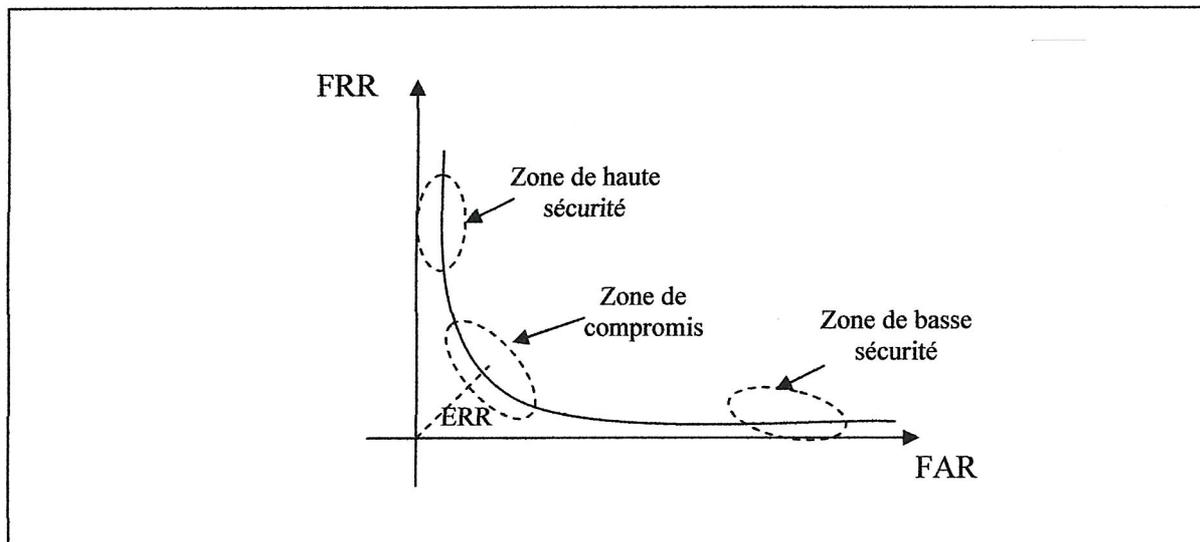


Figure III.2. Courbe ROC (Receiver Operating Characteristic).

- pour un système d'identification, la mesure de performance la plus utilisée est le taux d'identification. En cas d'erreur, il est nécessaire de tracer la courbe score cumulé (Cumulative Match Score) représentant la probabilité que le bon choix se trouve parmi les N premiers. Dans le cas d'une base de données contenant plusieurs modèles pour chaque individu, on calcule la précision (Precision) et le rappel (Recall),

- pour un système de vérification, la performance se mesure à son taux de faux rejets (False Rejection Rate: FRR) et celui de fausses acceptations (False Acceptation Rate:

FAR). Pratiquement on doit faire un compromis entre FRR et FAR et ceci en déterminant un seuil de décision. Ce dernier est déterminé par deux critères: l'erreur totale (TE) et le taux d'erreur égale (Equal Error Rate: EER). Pour le premier cas, il minimise l'erreur totale et pour le deuxième cas équilibre les deux taux (FAR et FRR). A cette fin, on trace pour chaque seuil, la courbe ROC (Receiver Operating Characteristic) qui présente FRR en fonction du FAR (figure III.2). L'erreur totale est donnée par :

$$ET = \frac{FA + FR}{V + I} \tag{III.1}$$

avec *FA* : nombre de fausses acceptations,
FR : nombre de faux rejets,
V : nombre d'accès clients,
I : nombre d'accès imposteurs.

Le taux de fausses acceptations et celui de faux rejets vérifient respectivement :

$$FAR = \frac{FA}{I} \tag{III.2.a}$$

$$FRR = \frac{FR}{V} \tag{III.2.b}$$

I.5 Exemples de caractéristiques biométriques

Nombreuses, sont les caractéristiques biométriques, mais nous avons choisi de donner un bref aperçu de quelques caractéristiques physiologiques et d'autres comportementales. Le tableau III.1 présente ces caractéristiques avec leur appareil d'acquisition.

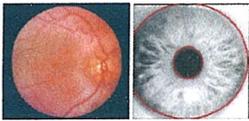
Caractéristique biométrique	Moyen d'acquisition	Caractéristique biométrique	Moyen d'acquisition
(a) 		(b) 	
(c) 		(d) 	
(e) 		(f) 	

Tableau III.1: Exemples de caractéristiques biométriques et leur appareil d'acquisition : (a) la rétine et l'iris, (b) les empreintes digitales, (c) la géométrie de la main, (d) le visage, (e) la signature et (f) la voix.

I.5.1 L'iris (Iris)

L'iris est la zone colorée visible entre le blanc de l'oeil et la pupille, son identification fut proposée en 1936 par l'ophtalmologiste Frank Branch [41]. En 1987, deux professeurs Aran et Leonard Flon déposèrent un brevet concernant cette technologie et en 1989 ils demandèrent l'aide du professeur Jhon Daughman pour réaliser un algorithme mettant au point cette technique. Cet algorithme a été breveté en 1994 et devenu la base de la reconnaissance de l'iris. Richard P. Wildes a également proposé en 1997 [42] un système d'identification et de vérification basé sur l'iris.

Les systèmes biométriques basés sur cette caractéristique sont [43] infaillibles compte tenu que l'iris est stable le long de la vie, distinct et sa reproduction artificielle est pratiquement impossible. Malgré qu'ils se prétendent aux vérifications 1:1 et à l'identification 1:N, ces systèmes sont coûteux, peu acceptés par les utilisateurs et nécessitent des équipements spécifiques. La banque « United Of Texas » était la première banque des Etats Unis qui a appliqué cette technologie aux guichets automatiques.

I.5.2 La rétine (Retina)

La rétine est un tissu nerveux tapissant le fond de l'oeil et c'est la disposition de ses veines qui permet la distinction entre les rétines d'un individu à l'autre et d'un oeil à l'autre d'une même personne. L'identification basée sur la rétine remonte à 1936 [41] avec les travaux du Dr. Carleton Simon et Dr. Isadore Goldstein.

Exploitant le fait que la rétine est unique, propre à chaque individu et stable le long de la vie [43], les systèmes biométriques qui l'utilisent sont très fiables mais coûteux, intrusifs et peu acceptés par les utilisateurs ce qui réserve cette technique aux applications qui exigent une haute sécurité et concernent la sûreté nationale (salle de coffre fort, site militaire,...) soit dans le mode 1:1 ou le mode 1:N.

I.5.3 La géométrie de la main (Hand geometry)

Les systèmes biométriques utilisant la géométrie de la main (longueur et largeur des doigts, largeur de la paume,...) sont les plus utilisés et les plus adaptés au contrôle d'accès puisqu'ils conviennent aux vérifications 1:1.

Cette caractéristique n'est pas un moyen biométrique fiable puisque [43] les systèmes peuvent être facilement trompés par de vrais jumeaux ou même par des personnes de la même famille, c'est pourquoi la fusion avec d'autres caractéristiques telles que l'empreinte de la paume [45] ou la géométrie des doigts est une alternative à la géométrie de la main.

I.5.4 Les empreintes digitales (Fingerprints)

Les empreintes digitales sont le relief cutané des bouts des doigts. Pour les voir, il suffit de couvrir les bouts des doigts d'une fine couche d'encre et de les imprimer sur une feuille de papier. Ces dernières ont été utilisées, à des fins d'identification, par les chinois pour signer les documents, il y a un millénaire, mais il fallait attendre l'anthropologue Francis Galton qui démontra, en 1882, la permanence et l'unicité des empreintes digitales (même pour les jumeaux monozygotes [43]) pour qu'elles soient adoptées par les services judiciaires (Scotland Yard dès 1901 et la préfecture de Paris dès 1903).

Cette caractéristique est fiable mais mal acceptée par les utilisateurs à cause de son association avec la criminologie. Parmi les applications des systèmes biométriques utilisant les empreintes digitales, on cite la personnalisation de voitures [43] proposée par Siemens et la matrice active [38] à empreintes digitales placée sur les ordinateurs portables de COMPAQ et TOSHIBA.

I.5.5 La voix (Voice)

En 1962, Lawrence Kersta établit que la voix de chaque personne est unique et qu'il est possible de la représentée graphiquement [41]. Cependant la voix peut changer le long de la vie à cause d'une maladie, de l'état émotionnel de la personne et même à cause du froid c'est pourquoi la reconnaissance vocale [43] relève à la fois de la biométrie physiologique et comportementale.

Comme un moyen de vérification 1:1, la voix est acceptée par les gens mais sa fiabilité est moyenne. Elle s'applique notamment pour contrôler le passage des frontières et pour sécuriser les opérations bancaires et commerciales en lignes,...

I.5.6 La signature (Signature)

Signer pour s'identifier est un acte quotidien, notamment pour les transactions vu que la signature manuscrite est propre à chaque individu. L'identification par la signature se fait en deux modes : statique ou dynamique; le premier utilise l'information géométrique de la signature, quant au deuxième, il est le plus riche en informations qui proviennent de la vitesse avec laquelle la signature est réalisée, la précision et la position spatiale du stylo à encre (inkpen), de la trajectoire de la signature, et de l'inclinaison,...

Ce moyen biométrique est accepté par les utilisateurs mais présente l'inconvénient de dépendre de l'état émotionnel, de la fatigue, de la santé du signataire et de la variabilité entre deux signatures d'une même personne.

I.5.7 Thermogramme facial (Facial thermogram)

Cette technique biométrique utilise la chaleur émise par le visage pour identifier une personne. L'image est acquise par une camera infrarouge indépendamment des conditions d'illumination, cependant elle est influencée par la pose, la température corporelle et l'état émotionnel de l'individu.

I.5.8 Le visage (Face)

L'une des habilités les plus surprenantes du système visuel humain est celle d'identifier les personnes à partir de leur visage, indépendamment des expressions faciales et des changements dus à l'âge et aux styles de coiffures. Cette capacité a poussé les chercheurs, il y a plus de 30 ans à introduire des systèmes automatiques de reconnaissance de visages basés soit sur les relations géométriques entre les éléments caractéristiques du visage (yeux, nez, bouche,...), soit sur le visage entier.

Ce moyen biométrique, dédié à la vérification et/ou à l'identification, présente l'avantage, comparé à d'autres caractéristiques biométriques, d'être non intrusif et acceptable par les utilisateurs, seulement, il dépend fortement des conditions d'éclairage, de la position de la tête, des expressions faciales, des occultations partielles et du temps. En effet, pour un même système de reconnaissance de visages, le taux de reconnaissance peut chuter de 50% [43].

I.6 Multimodalité

Certes que les caractéristiques biométriques sont un moyen d'identification efficace et précis, mais elles dépendent énormément des conditions d'acquisition (éclairage pour les visages, bruits pour la voix, type de travail pour les empreintes digitales,...), des capteurs utilisés (scanner optique, scanner thermique, camera,...), de l'état de santé et de l'état émotionnel des individus et aussi des méthodes utilisées lors de l'extraction de caractéristiques pour former les vecteurs d'attributs.

D'autres part, certains systèmes biométriques unimodaux sont coûteux et mal acceptés par les utilisateurs tels que les systèmes de reconnaissance de l'iris et de la rétine, c'est pourquoi la multimodalité a été utilisée pour pallier ces problèmes, parmi d'autres, tels que la non disponibilité de certaines caractéristiques à cause d'un accident ou d'une maladie. Ainsi un système biométrique multimodal est plus performant (précision, efficacité), résistant à l'imposture (impossible de mimer plusieurs caractéristiques biométriques) et robuste puisqu'il compense une modalité défaillante par une ou d'autres modalités.

I.6.1 Différentes formes de multimodalité

Le sens le plus classique du terme multimodalité est la fusion de plusieurs caractéristiques biométriques telles que le visage, les empreintes digitales et la géométrie de la main [44] en plus de la parole et la signature manuscrite dans le projet BIOMET [39], mais il y a d'autres formes de fusion [43][44] :

1. **Systèmes multiples d'acquisition** qui utilisent différents capteurs pour acquérir une même caractéristique biométrique,
2. **Mesures multiples d'une même unité biométrique** telle que la reconnaissance d'une personne en utilisant les empreintes digitales de ses 10 doigts ou de ses deux iris,
3. **Instances multiples d'une mesure** qui consiste à répéter l'acquisition d'une même caractéristique avec le même capteur telle que prendre plusieurs empreintes digitales d'un doigt de la même personne,
4. **Algorithmes multiples**: faire la reconnaissance à partir d'une seule caractéristique biométrique mais avec différents algorithmes de reconnaissance.

Toutes ces techniques de fusion peuvent être évaluées selon le coût matériel et/ou logiciel, l'apport d'informations par rapport à un système unimodal et la gêne pour les utilisateurs [43].

I.6.2 Fusion d'informations dans les systèmes biométriques multimodaux

La fusion des informations dans les systèmes bimodaux ou multimodaux peut s'effectuer lors de l'une de ces trois étapes [44]:

- **extraction de caractéristiques (Feature extraction)**: dans cette étape, on concatène plusieurs vecteurs obtenus de différentes caractéristiques pour former un seul vecteur hybride mais ayant une grande dimension, ce dernier sera l'entrée du classificateur,
- **étape de comparaison (matching)**: le résultat final est obtenu en fusionnant les résultats de similarité entre chaque vecteur et son modèle,
- **étape de classification (decision)**: les vecteurs caractéristiques issus de plusieurs capteurs sont classifiés séparément pour retourner l'acceptation ou le rejet dans le mode de vérification et l'identité dans le mode d'identification.

La fusion, dans ce cas, revient à combiner les résultats des classificateurs utilisés, elle peut se faire soit [43] : **i)** au niveau abstrait en fusionnant, par exemple, à l'aide d'un vote à la majorité ou de type K parmi M (M est le nombre de vecteurs caractéristiques à classifier), **ii)** au niveau des mesures à l'aide des moyennes simples ou pondérées des probabilités d'appartenance aux classes ou **iii)** au niveau des rangs où les résultats sont affinés avec une mesure de confiance et la fusion est réalisée à l'aide des règles telles que: le maximum, le minimum et le médian.

I.7 Bases de données multimodales

Réaliser une base de données multimodale consiste à collecter, pour chaque personne participant à cette base, différentes modalités. Parmi les bases multimodales existantes, on compte: **M2VTS** et **XM2VTS** [43] qui regroupent à la fois des séquences vidéo et sons synchronisés de 37 et 295 individus respectivement, la base **BIOMET** [39] créée en 2002 avec cinq modalités, de près de 100 personnes, à savoir: visage, parole, signature, empreintes digitales et forme de la main. Elle peut être utilisée, à la fois, pour des systèmes biométriques unimodaux et multimodaux et une autre base [44] contenant trois modalités de 50 personnes (visage, empreinte et géométrie de la main) réalisée à des fins de vérification.

II Reconnaissance de Visages

II.1 Historique de la reconnaissance de visages

Historiquement, la reconnaissance de visages (RdV) [60] remonte à 1888 avec les travaux de Galton à des fins d'identification de personnes à partir de leur vues de profil et avec les études psychologiques de Brunier et Togiuni en 1954 qui portaient sur l'étude du système visuel humain, mais les recherches concernant la reconnaissance automatique de visages n'ont vu le jour qu'en 1966 avec le premier système semi automatique de Bledsoe. Il consistait à reconnaître les visages après avoir marqué manuellement des points caractéristiques sur des visages photographiés (les coins des yeux, les commissures de la bouche, le nez, le menton,...), suivi du système proposé par Kelley en 1970 utilisant aussi des mesures géométriques. Le premier système entièrement automatique, proposé par Kanade date de 1973 [102], il consistait à extraire automatiquement des points caractéristiques à partir d'une seule image. Tous ces systèmes comparaient des mesures géométriques (distances, rapports de distances,...) d'un visage inconnu avec celles stockées.

En 1991, la publication de l'article « Eigenfaces for Recognition » [50] fut un tournant dans le domaine de la reconnaissance de visages, depuis les articles et les publications ne cessent d'apparaître en s'inspirant des eigenfaces. Le fruit de toutes ces recherches était les produits commerciaux de reconnaissance faciale motivés par le programme FERET (Face Recognition Technologie) organisé par le ministère de la défense américaine en 1994 et terminé en 1996. Ce programme avait pour but l'évaluation des algorithmes de RdV sur une grande base de visages (14126 images de 1199 personnes) acquise à un intervalle d'un à deux ans.

Parmi les producteurs de ces systèmes, on cite : FaceIT from Visionics, Viisage Technologie, Biometric System, IN, Face Snape Recorder, etc. Leur commercialisation envahit le marché des systèmes biométriques avec, selon International Biometric Group [38], des ventes qui dépassent celles des systèmes basés sur la voix. Ces systèmes interviennent dans plusieurs applications telles que:

- **la sécurité d'informations** : sécurité de bases de données, cryptage de données, accès à l'Internet,

- **la répression et la surveillance** : poursuite de suspects et investigation, contrôle d'accès,
- **le divertissement** : jeux vidéo, interaction homme-robot, interaction homme-ordinateur,
- **les cartes (smart cards)** : carte d'identité, passeport, registration des électeurs.

Parmi les applications les plus populaires : le logiciel Mandrake mis en service le 14 octobre 1998 à Borough de New Ham de Londres visant à diminuer le nombre de crimes et de délits (il a abouti à une diminution de 10% en 6 mois), la vidéo surveillance utilisée dans le stade où se déroulait la finale de football qui a eu lieu à Tampa en Floride en janvier 2001, le système de RdV en Ouganda en mai 2001 lors de l'élection présidentielle, les systèmes utilisés en Afrique du sud dans une grande chaîne de distribution de paiement et dans les aéroports pour le contrôle de passeports,...

II.2 Visage humain comme stimuli

Le visage est le noyau des systèmes de reconnaissance de visages, c'est pourquoi il mérite un regard d'intérêt de notre part. Ce dernier est la face antérieure de la tête allant du front jusqu'au menton et d'une oreille à l'autre, il se constitue d'un regroupement d'éléments distincts (yeux, nez, bouche,...) disposés selon une organisation géométrique similaire (les yeux sur les cotés, le nez au milieu, la bouche au dessous du nez,...) et il est riche en informations sociales.

Malgré qu'on possède les mêmes éléments faciaux, ces derniers diffèrent d'une personne à une autre: les yeux peuvent être ronds ou minces, le nez peut être grand ou petit, pointu ou émoussé, la bouche peut être grande ou petite, la couleur de peau peut être claire ou foncée,...., l'état émotionnel, aussi, influence sur les éléments faciaux: une personne étonnée peut avoir la bouche ouverte et faire de grands yeux et une personne fâchée peut avoir un regard sérieux et la bouche fermée,... tous ce qu'on vient de dire rend le visage un stimuli visuel de classe à part.

II.3 Psychologie / Neurosciences et reconnaissance de visages

Avec toutes les variances inter et intra visages, il nous suffit un clin d'oeil pour distinguer le sexe, l'état émotionnel et l'identité d'une personne à partir de son visage même après plusieurs années de séparation et avec les changements que peut subir le visage tels que l'expression faciale, l'âge, les cheveux, l'usage de produits cosmétiques et les accessoires,... Derrière cette grande capacité à identifier efficacement et rapidement les visages, résident deux hypothèses: **i)** la RdV est un processus spécifique qui diffère de la reconnaissance des autres objets et **ii)** il y a une aire spécifique du cerveau dédiée à la RdV. Ces hypothèses sont soutenues par les arguments suivants [60]:

- La RdV est plus facile et plus rapide que la reconnaissance des autres objets,
- Les patients atteints de prosopagnosie souffrent d'une altération des facultés en RdV indépendamment d'une détérioration des facultés en reconnaissance des objets (ils ne reconnaissent ni les visages de leurs proches, ni les siens sur une glace), cependant ces mêmes malades reconnaissent si l'objet est un visage ou non et distinguent souvent ses différentes parties mais ne peuvent pas l'identifier. D'autres patients atteints d'agnosie des objets visuels n'éprouvent aucune difficulté à reconnaître les visages.

Selon les études psychologiques, plusieurs facteurs contribuent plus au moins à performer le système visuel humain, parmi lesquels on cite [60]:

- **les expressions faciales:** la RdV est indépendante de la reconnaissance d'expressions faciales, puisque quelques patients prosopagnosiques reconnaissent les expressions faciales d'un visage mais ont du mal à l'identifier par contre, les patients atteints d'un syndrome organique du cerveau (Organic Brain Syndrome) souffrent d'une pauvre analyse d'expressions faciales mais reconnaissent parfaitement les visages,
- **le mouvement:** les visages célèbres et familiers sont plus faciles à identifier, à partir des séquences dynamiques, que des photographies statiques, cependant le mouvement n'apporte rien pour les visages non familiers,
- **le changement d'éclairage:** un bon éclairage est approprié pour mieux reconnaître les visages puisqu'il est difficile de les identifier à partir des photographies négatives,
- **la distinction:** les visages distincts sont bien mémorisés et rapidement reconnus que les visages typiques,
- **les caractéristiques faciales significatives:** les yeux, la bouche, les cheveux, le contour de la tête sont importants pour percevoir et se souvenir des visages. La performance de reconnaître un visage s'améliore aussi avec d'autres facteurs qui s'ajoutent aux premiers à savoir la beauté, l'attrance et/ou la sympathie.
- **l'analyse globale et par composantes (Holistic analysis, feature analysis):** toutes les deux sont cruciales pour la perception et la RdV ; l'analyse globale est suivie d'une analyse par composantes (yeux, nez, bouche,...) pour affinement. Cependant en présence de caractéristiques marquées telles qu'un long nez ou de grandes oreilles, l'identification peut s'en dépasser des informations globales.

La nécessité d'une analyse globale/ configurale (positions et distances relatives) pour la RdV a été démontrée par l'expérience de Thompson «Thatcher illusion», réalisée en 1980: les yeux et la bouche de Margaret Thatcher étaient inversés et placés sur le visage (figure III.3(c)), ce dernier apparaît grotesque tandis que son image inversée semble normale (figure III.3(d)).

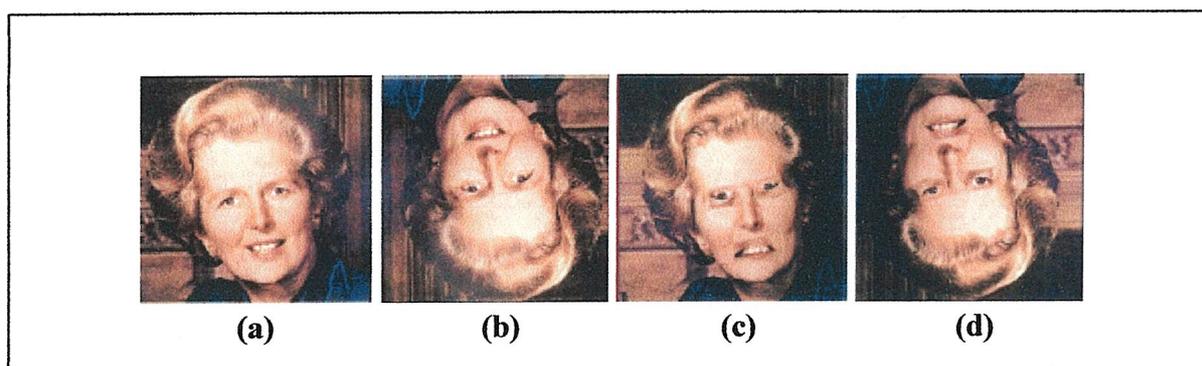


Figure III.3. Illusion de Thatcher: (a) visage normal, (b) visage inversé, (c) éléments du visage inversés et (d) visage de la figure(c) inversé.

II.4 Système de reconnaissance automatique de visages

II.4.1 Définition

Un système automatique de reconnaissance de visages est un système biométrique utilisant le visage à des fins d'identification et/ou de vérification de personnes à partir des

images de visages fixes [50][58] ou de séquences vidéo [46] en comparant les caractéristiques d'un individu avec celles stockées dans une base de référence. Il a l'avantage, comparé à d'autres systèmes biométriques, d'être non intrusif, effectuant l'identification d'une façon naturelle et acceptable par les utilisateurs. Il comporte essentiellement, après l'acquisition des images, les étapes illustrées en figure III.4:

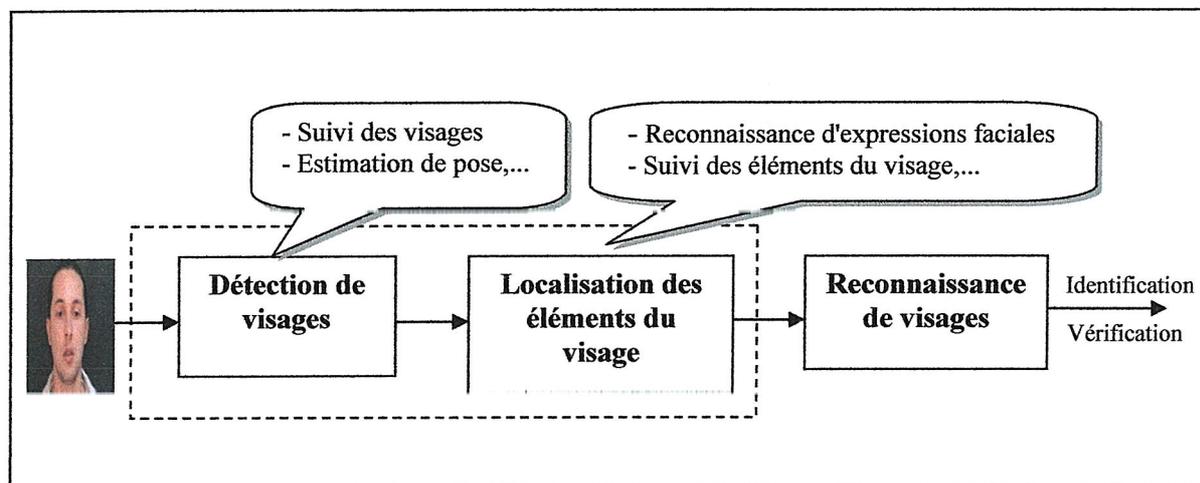


Figure III.4. Système de reconnaissance automatique de visages.

Ces étapes peuvent être combinées ou utilisées séparément pour d'autres applications (figure III.4). La détection et la localisation peuvent s'effectuer simultanément, quant à la localisation des éléments faciaux, elle peut servir à la reconnaissance comme à la détection. Soit pour une application ou une autre, toutes ces étapes doivent être réalisées soigneusement parce que la performance des applications dans lesquelles interviennent en dépend beaucoup: le suivi d'un visage nécessite sa détection et la reconnaissance d'expression faciale ne peut être sans la localisation des éléments faciaux.

II.4.2 Reconnaissance de visages à partir des séquences d'images vidéo

La reconnaissance de visages à partir des séquences vidéo est une extension de la reconnaissance à partir des images fixes. Elle est privilégiée à cause du mouvement qui aide à mieux reconnaître les visages familiers si leurs images sont négatives, inversées ou seuillées (Threshold). Le mouvement aide également à mieux identifier les visages animés d'un ensemble que ces mêmes visages aléatoirement arrangés. Cependant, la RdV à partir de séquences vidéo souffre, en plus des variations d'illumination, de la pose de la tête, des occultations partielles et de la non coopération des personnes, de la petite taille des images de visages. Cette taille rend difficile, non seulement, la tâche de reconnaissance mais influe également sur la précision de segmentation et de localisation des éléments caractéristiques du visage.

II.4.3 Détection et localisation de visages

La détection de visages est la première étape dans plusieurs systèmes automatiques : la RdV, le suivi des éléments faciaux, la reconnaissance des expressions faciales, ... Détecter des visages revient à vérifier la présence ou non d'un ou de plusieurs visages dans une image et de retourner la position approximative du visage détecté s'il existe. Plusieurs facteurs rendent cette tâche difficile à savoir les problèmes de pose (vue de face, profil, ...), d'expression faciale (visage neutre, souriant, en colère, ...), d'occultations partielles

(lunette, chapeau,...) et les conditions d'acquisition des images (éclairage, direction des sources de lumière,...). La littérature [73] reporte plus de 150 méthodes de détection de visages, qui peuvent être regroupées en quatre catégories:

- Méthodes basées sur les connaissances,
- Méthodes basées sur l'extraction de caractéristiques invariantes,
- Méthodes basées sur l'appariement de gabarit,
- Méthodes basées sur l'apparence.

Quant à la localisation de visages (figure III.5 (a)), elle est un cas particulier de la détection; pour lequel on suppose que l'image d'entrée ne contient qu'un seul visage et on cherche sa position.

II.4.4 Détection des éléments caractéristiques du visage

Elle vise à détecter la présence et après la localisation des différents éléments du visage tels que: les yeux et les sourcils, le nez, les narines, la bouche, les commissures,... en supposant l'unicité du visage dans l'image à traiter. Cette partie sera détaillée dans le paragraphe II.8 de ce chapitre.

II.4.5 Suivi de visages et d'éléments faciaux

A partir d'une séquence vidéo, le suivi de visages consiste à estimer la localisation et l'orientation d'un visage en temps réel. Les méthodes réalisant cette application peuvent être répertoriées en 3 catégories [60]: **i) le suivi de la tête** qui revient à suivre un objet rigide en rotation et en translation, **ii) le suivi des éléments caractéristiques du visage**; il consiste à suivre des déformations non rigides dues à la parole, aux expressions faciales et aux contractions et relaxations des muscles en utilisant des limites ou des régions contenant les éléments du visage. Le suivi de contours (feature boundary), par exemple, vise à suivre les mouvements et à délimiter la forme des éléments du visage tels que les lèvres et la bouche (figure III.5 (b) et figure III.5 (c)) et **iii) le suivi total (complete tracking)** qui combine les deux catégories précédentes.

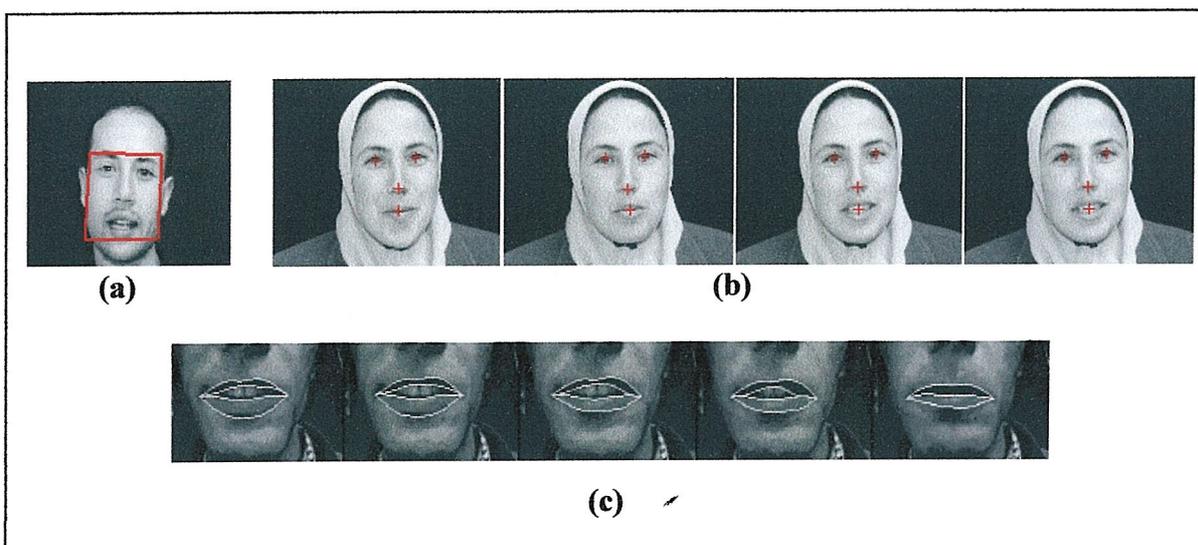


Figure III.5. (a) localisation d'un visage, (b) localisation des éléments du visage à partir d'une séquence d'images et (c) détection et suivi des contours labiaux d'une séquence d'images [67].

II.4.7 Reconnaissance d'expression faciale

Elle a pour but l'identification de l'état affectif d'une personne pour renvoyer l'une des expressions innées correspondantes aux sept émotions suivantes : la neutralité, la joie, la tristesse, la surprise, la peur, la colère et le dégoût.

II.5 Approches de la reconnaissance de visages

Plus de 3 décennies de recherche ont conduit à de nombreux systèmes de RdV basés sur différentes méthodes mais que l'on peut regrouper en trois catégories [60] à savoir:

II.5.1 Approches basées sur les éléments caractéristiques du visage (Feature-based approaches)

Les approches géométriques étaient les premières approches utilisées en RdV. Elles consistaient à localiser différents points caractéristiques et calculer ensuite des mesures géométriques entre eux (longueur, largeur, distances, rapports de distances, angles,...) [46][65][74][102]. Cela à partir des vues de faces [46], de profil [89], où les deux [88]. Parmi lesquelles, nous comptons:

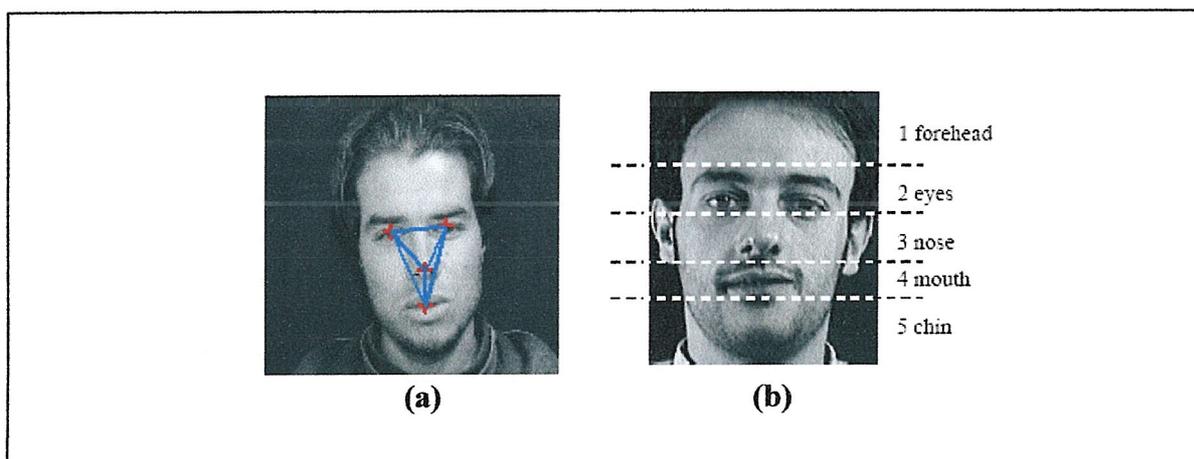


Figure III.6. (a) Exemple de six distances entre les éléments du visage, et (b) bandes faciales utilisées dans [103].

Le système de Kanade (1973) [102] qui portait sur l'utilisation de 16 mesures géométriques et donnait un taux de reconnaissance de 75% sur une base de 20 personnes contenant 40 images.

Brunelli et Poggio (1993) [74] réalisaient un système de RdV, après l'extraction automatique de 35 distances. A l'aide du classificateur de Bayes, ils arrivaient à un taux de reconnaissance de 90% sur une base contenant 188 images de 47 personnes.

En 1996, Cox et al. [65] ont proposé un système basé sur un mélange de distances (Mixture Distance). Ils avaient extrait manuellement 30 distances de 685 individus et arrivaient à un taux de reconnaissance de 95%.

Harmon et al. [89] ont utilisé 17 caractéristiques faciales à partir des vues de profil. Leur méthode arrivait à reconnaître 96% d'une base de 112 personnes.

Récemment, deux méthodes géométriques utilisant un PMC ont été proposées. L'une pour la vérification [79] et l'autre pour l'identification de visages [96] en utilisant respectivement six et huit distances entre les yeux, le nez et la bouche (voir la figure III.6(a)).

Une autre méthode, très importante, basée sur les Modèles de Markov Cachés (HMM) a été proposée en 1994 par Samaria [103]. Après avoir partitionner le visage en bandes (figure III.6(b)), la méthode HMM monodimensionnel (One-Dimensional HMMs) a donné, sur la base ORL, un taux de 87% et la méthode HMM pseudo 2D (Pseudo Two-Dimensional HMMs) un taux de 95%.

Les capacités de ces approches sont limitées. Cela est du essentiellement à l'occultation de certaines caractéristiques faciales (vue de profil, port de lunette, moustache, cache-nez, ...) et aux algorithmes de localisation des éléments caractéristiques de visages qui sont jusqu'à présent peu performants.

II.5.2 Approches globales (Holistic approaches)

Elles utilisent l'information apportée par le visage entier sans segmentation de ses parties. Pour les appliquer, les visages de la base de données représentés par des matrices de n lignes et de m colonnes chacun, doivent être transformés en vecteurs en concaténant les lignes (ou les colonnes) de ces matrices afin d'obtenir des vecteurs de dimensions $(n \times m, 1)$ qui seront fournis au classificateur pour la reconnaissance.

Ces méthodes souffrent de la malédiction de la dimension (Curse of dimensionality), c'est pourquoi plusieurs méthodes de réduction de dimensions linéaires et non linéaires (ACP, ADL, ACI, ACP à noyau, ...) ont été adoptées dans les systèmes de reconnaissance de visages menant à différentes approches parmi lesquelles on cite: eigentaces [51], Fishertaces [57], Subspace LDA [58][66], ICA [84], Evolution Pursuit [85], etc. D'autres représentation ont été aussi proposées, telles que PDBNN (Probabilistic Decision Based Neural Network) [86], Feature line [47], Tensorfaces [82] et Laplacianfaces [56], ... mais récemment la tendance est d'appliquer les techniques d'extraction de caractéristiques bidimensionnelles soit la version linéaires ou non linéaires à savoir l'ACP2D [52], l'ACP2D bilatérale et l'ACP2D à noyau [83], l'ADF2D [54][77][87], l'ADF2D bilatérale et l'ADF2D à noyau [91] et l'ADL2D bilinéaire [55],...

Cette catégorie dépend énormément des conditions d'éclairage, des expressions faciales, de la pose de la tête, des occultations partielles et de l'échelle des visages, ... Nous allons détailler les deux méthodes eigenfaces et Fisherfaces.

II.5.2.1 Eigenfaces

Sirovitch et Kirby, en 1987, étaient les premiers [51][50][60] à utiliser l'ACP pour coder efficacement les images de visages avec ce qu'ils appelaient "eigenpictures". Leur travail a motivé Turk et Pentland en 1991 [51][50] à introduire leur méthode " eigenfaces " pour la détection et la reconnaissance de visages et depuis eigenfaces est devenue l'approche la plus utilisée dans ce domaine jusqu'à la considérer une référence par rapport à laquelle les chercheurs comparent leur approches, et une source d'inspiration pour créer de nouvelles méthodes. Les eigenfaces sont les vecteurs propres de la matrice de covariance des images de visages et ils ressemblent aux visages, c'est pourquoi Turk et Pentland l'ont appelés eigenfaces.

a) Algorithme

Nous avons résumé et simplifié la méthode des eigenfaces proposées dans [51][50] par l'algorithme III.1 en utilisant une formulation matricielle. Si l'on dispose d'une base de M images de visages en niveau de gris avec plus d'une image par personne, les matrices-visages $X_j (n \times m)$ doivent être transformées en vecteurs $\Gamma_j (n \times m, 1)$.

Algorithme III.1: Eigenfaces**Entrées:** vecteurs des images Γ_j avec $j = 1, \dots, M$ **Sorties:** eigenfaces et composantes principales

1. Former un ensemble de M vecteurs images: $P = [\Gamma_1 \Gamma_2 \dots \Gamma_M]$,
2. Calculer le visage moyen ψ selon: $\psi = \frac{1}{M} \sum_{j=1}^M \Gamma_j$,
3. Soustraire à chaque vecteur de visage le visage moyen: $\Phi_j = \Gamma_j - \psi$,
Rassembler ces vecteurs dans la matrice A telle que: $A = [\Phi_1 \Phi_2 \dots \Phi_M]$,
4. Calculer les valeurs propres λ_j et les vecteurs propres v_j de la matrice $L = A^T \cdot A$ de taille $(M \times M)$,
5. Ordonner les valeurs propres en ordre décroissant et retenir d vecteurs propres correspondant à d valeurs propres les plus grandes
Rassembler ces d vecteurs propres dans une matrice: $V = [v_1 v_2 \dots v_d]$,
6. Calculer les vecteurs propres de la matrice de covariance $C = A \cdot A^T$ selon: $U = A \cdot V$,
7. Projeter les vecteurs de visages dans l'espace de visages: $W = U^T \cdot A$
8. (optionnel) Reconstruire les images de visages: $\tilde{A} = U \cdot W$ en ajoutant à chaque colonne le visage moyen ψ .

Pour appliquer cet algorithme, nous avons utilisé la base de visages B-LAIG présentée et détaillée dans [69]. Cette application nous a permis d'obtenir des eigenfaces (visages propres) et de reconstruire les images de visages en fonction de différents nombres de vecteurs propres retenus. La figure III.9 illustre des exemples de cette application.

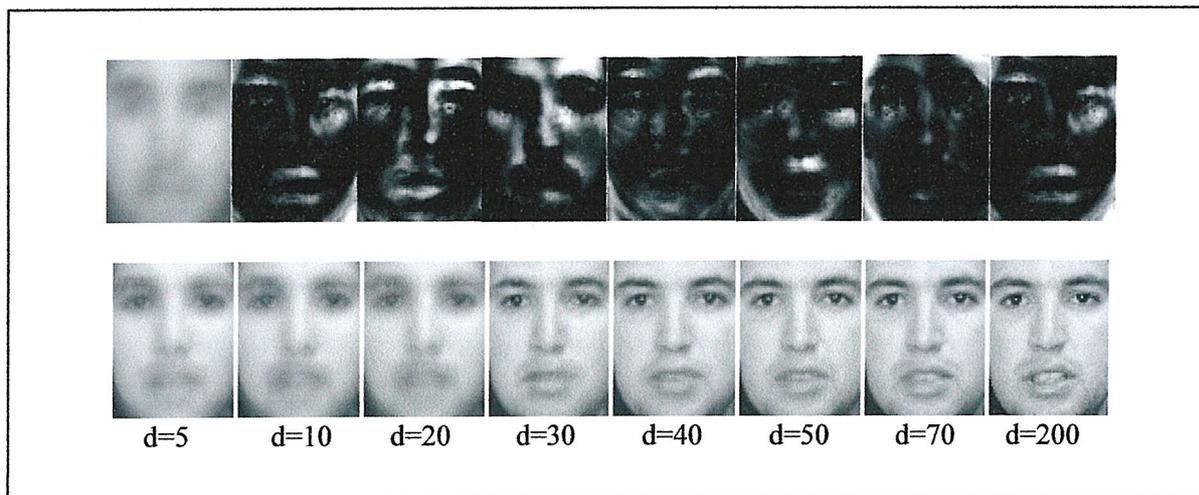


Figure III.9. 1^{ère} ligne: image moyenne (à gauche) et exemples de visages propres de la base B-LAIG, 2^{ème} ligne : reconstruction des images de visages en fonction des vecteurs propres.

b) Limitations

Malgré son efficacité, la méthode eigenfaces souffre des problèmes suivants :

- Les trois premières composantes principales des eigenfaces retiennent les variations dues à l'éclairage et aux expressions faciales [57],
- Eigenfaces est plus adaptée à la représentation de visages qu'à leur classification [66],
- Eigenfaces dépend du nombre de vecteurs propres retenus,
- La matrice de covariance totale est très coûteuse en temps de calcul et en espace mémoire.

Pour remédier au dernier problème, Turk et Pentland [51][50] ont proposé une technique de calcul (étape 4,5 et 6 de l'algorithme III.1) permettant le passage d'une matrice de covariance de taille $(n \times m, n \times m)$ à une autre de taille $(M \times M)$ où $M \ll (n \times m)$ est le nombre des images de visages de l'ensemble d'apprentissage, cette technique donne un gain aussi bien en temps de calcul qu'en espace mémoire.

c) Choix du nombre de vecteurs propres à retenir

Le nombre de vecteurs propres associés aux plus grandes valeurs propres à retenir est un grand défaut de cette technique. Pour le choisir, les chercheurs ont adopté différentes solutions:

- Pour un ensemble de 115 images, Sirovitch et Kirby ont trouvé que 40 eigenpictures sont suffisantes pour représenter efficacement cet ensemble,
- Turk et Pentland [51][50] l'ont choisi heuristiquement. Pour leurs tests, sur une base de 16 individus, $d = 7$ valeurs propres,
- Moghaddam [53] a préservé, pour comparer différentes approches de RdV 20 vecteurs propres en justifiant son choix par une erreur de reconstruction raisonnable (0.0012) et un taux de reconnaissance 80% obtenu par eigenfaces sur une base de 1829 images,
- Zhao et al. [58] ont retenu 300 vecteurs propres pour une base de 1038 images après avoir observé que pour un nombre très élevé, les eigenfaces ne représentent pas des visages, donc leur choix était basé sur l'allure des eigenfaces au lieu des valeurs propres.

Pratiquement, le choix de d peut se faire: **i)** en fixant un certain pourcentage (par exemple 5%) tel que [66] :

$$\frac{\sum_{i=d+1}^n \lambda_i}{\sum_{i=1}^n \lambda_i} < P \quad \text{avec } n \text{ le nombre total des valeurs propres} \quad \text{(III.3)}$$

Ou **ii)** on se basant sur le fait que la $i^{\text{ème}}$ valeur propre est égale à la variance le long de la $i^{\text{ème}}$ composante principale et on cherche donc d le long du spectre des valeurs propres (eigenvalue spectrum) (figure III.8) juste avant qu'il s'annule [53][10].

Donc, les d premiers vecteurs propres correspondant à d valeurs propres les plus grandes est un paramètre critique sur lequel dépend la performance du système de RdV (temps de calcul et taux de reconnaissance) et aussi un critère d'évaluation [53][61] de méthodes basées sur l'ACP. Son choix dépend des contraintes d'application liées au temps de calcul et au nombre des images de visages disponibles [53] et aussi de la qualité des images de visages reconstruites [52].

La figure III.8 illustre les valeurs propres obtenues en appliquant eigenfaces sur la base de visages B-LAIG (nous avons utilisé 200 images de visages).

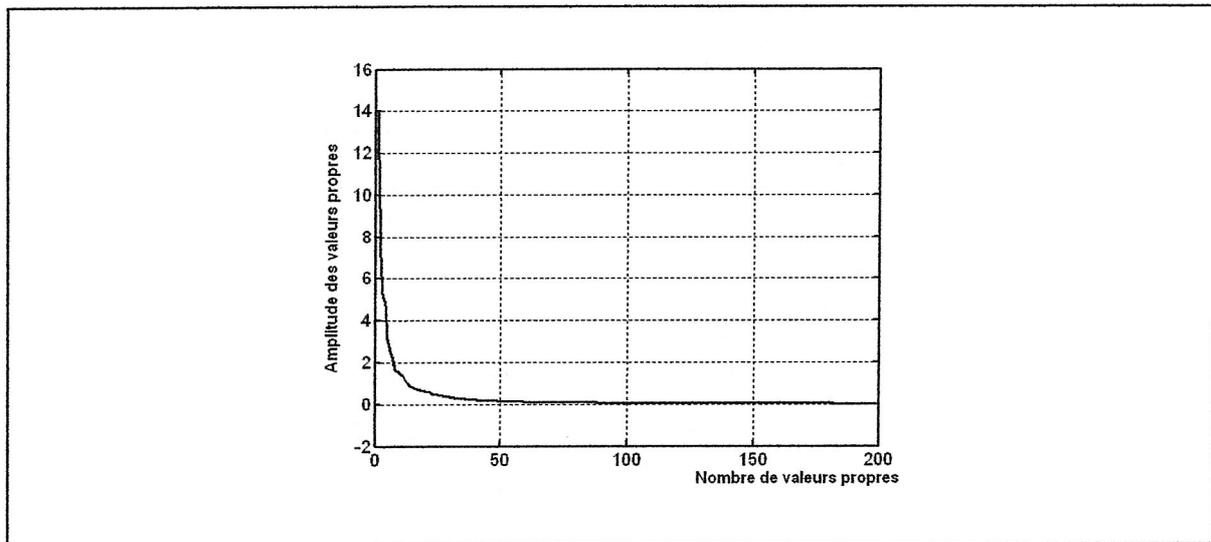


Figure III.8. Décroissance des valeurs propres obtenues par eigenfaces.

II.5.2.2 Analyse discriminante linéaire (ADL)

Pour la reconnaissance de visages, l'ADL attire aussi les chercheurs motivés par les arguments suivants [58][59]:

- l'ADL possède un pouvoir discriminant qui dépasse celui de l'ACP,
- l'ADL est un bon classificateur pour les données linéairement séparables,
- l'ADL est facile à implémenter.

mais malgré ses avantages, elle souffre du problème de la singularité de la matrice de covariance intra-classe S_w du au fait que le nombre de visages disponibles pour l'apprentissage est très inférieur comparé à la dimension des vecteurs de visages.

Pour corriger ce problème, Swets et Weng [66] ont proposé «la projection de discriminante Karhunen-Loève: Discriminant Karhunen-Loève Projection», Belhumeur et al. [57] ont introduit une méthode dite Fisherfaces, une dérivée de la discrimination linéaire de Fisher, Zhao et al. [58][59] ont remplacé la matrice de covariance intra-classe S_w par $(S_w + \Delta \cdot I)$ où Δ est une petite constante positive assurant que $(S_w + \Delta \cdot I)$ est strictement définie positive. Cette technique corrige, non seulement, le problème de la singularité de S_w mais permet également d'appliquer l'ADL en présence d'un seul exemple par classe.

D'autres méthodes ont été proposées pour pallier ce problème à savoir l'ADL Pseudo-Inverse [106], Regularized LDA [107] et une SVD généralisée [108]. Pour augmenter le pouvoir discriminant après avoir appliqué l'ACP suivie de l'ADL, d'autres techniques ont été introduites, parmi lesquelles on cite Direct-LDA [109] et Null-Space LDA [110]. Récemment, de nouvelles approches utilisant une représentation bidimensionnelle basée sur l'ADL ont été introduites [54][77][87][91], elles ont l'avantage d'être plus performantes que l'ADL ne souffrent pas du problème de la singularité et diminuent le temps de calcul.

Après avoir appliqué l'ADL, Zhao et al.[58][59] ont conclu qu'elle ne performe pas bien pour les 3 cas suivants: i) les images de test appartiennent à des personnes qui n'ont pas

participé à l'apprentissage, **ii)** les classes d'apprentissage contiennent des exemples largement différents, et **iii)** la présence de différents arrières plans.

II.5.2.3 Fisherfaces

Fisherfaces est une méthode dérivée de la discrimination linéaire de Fisher, elle a été proposée par Belhumeur et al.[57] à des fins de RdV en présence des variations d'illumination et d'expressions faciales. Son principe consiste à éviter le problème de la singularité de la matrice de covariance intra-classes en proposant une alternative au critère de Fisher. Cela en appliquant une ACP pour réduire la dimension de l'espace caractéristique à $(M - C)$ ensuite une ADL classique pour réduire la dimension à $(C - 1)$ où M est le nombre d'exemples d'apprentissage et C le nombre de classes.

Algorithme

L'algorithme suivant explique les étapes de la méthode Fisherfaces.

Algorithme III.2: Fisherfaces

Entrées: vecteurs des images Γ_j avec $j = 1, \dots, M$

Sorties: Fisherfaces et composantes principales

1. Former un ensemble de M vecteurs images centrés et répartis en C classes:

$$P = [\Gamma_1^{(1)} \Gamma_2^{(1)} \dots \Gamma_{n_1}^{(1)} \Gamma_1^{(2)} \Gamma_2^{(2)} \dots \Gamma_{n_2}^{(2)} \dots \Gamma_1^{(C)} \Gamma_2^{(C)} \dots \Gamma_{n_C}^{(C)}]$$

2. Appliquer l'ACP pour trouver la matrice de projection W_{PCA} telle que:

$$W_{PCA} = \arg \max_W |W^T S_T W| \quad \text{avec} \quad S_T = \sum_{j=1}^M (X_j - \bar{X})(X_j - \bar{X})^T$$

3. Appliquer l'ADL pour trouver W_{FLD} : $W_{FLD} = \arg \max_W \frac{|W^T W_{PCA}^T S_B W_{PCA} W|}{|W^T W_{PCA}^T S_W W_{PCA} W|}$

4. Calculer la matrice de projection optimale: $W_{opt}^T = W_{FLD}^T \cdot W_{PCA}^T$

5. Projeter les vecteurs de visages dans le nouvel espace de visages: $Y = W_{opt}^T \cdot X$
-

L'application de Fisherfaces [57] sur deux bases de visages à savoir la base Harvard et la base Yale, montre sa supériorité sur la méthode de corrélation et des eigenfaces avec et sans les trois premières composantes principales.

II.5.3 Approches hybrides (Hybrid approaches)

Imitant le système visuel humain qui utilise, pour la RdV, les informations globales et les composantes locales, des systèmes ont porté sur la fusion des deux méthodes précédentes pour améliorer la performance des systèmes automatiques de reconnaissance de visages. Au sommet de ces méthodes, « Modular eigenspaces » proposée par Pentland et al. en 1996 [48] qui consistait à combiner les méthodes eigenface et eigenfeatures. Testée sur une base de 45 personnes avec deux vues pour chacun, le visage neutre pour l'apprentissage et le visage

souriant pour le test, « Modular eigenspaces » donnait un taux de 98% dépassant ainsi celui des eigenfaces et des eigenfeatures.

Très récemment, Visani et al.[55] ont proposé leur méthode nommée «Analyse Discriminante Bilinéaire Modulaire (MBDA) » où l'ADL bilinéaire est appliquée sur trois régions du visage : une région contient tout le visage, la deuxième contient les sourcils, les yeux et le nez et la troisième ne contient que les sourcils et les yeux. Les auteurs reportent une supériorité de leur méthode sur Modular eigenspaces particulièrement en présence des variations dans les expressions faciales et la pose de la tête.

A notre connaissance, Brunelli et Poggio.93 [74] sont les seuls qui ont comparé, sur une même base de visages, une méthode géométrique et une méthode hybride (fusion des résultats obtenus). Cette comparaison a démontré la supériorité de la méthode hybride avec un taux de reconnaissance de 100%.

II.6 Problèmes d'illumination et de pose

Les expressions faciales (figure III.9), les occultations partielles, le style de coiffure, les accessoires, les produits cosmétiques, le vieillissement,... sont des facteurs qui influencent sur les performances des systèmes de RdV. Les chercheurs tentent de proposer des méthodes assez robustes en les testant en présence de différents problèmes, notamment les expression faciales, l'éclairage, la pose de la tête et le temps [57][49][111][52].



Figure III.9. *L'effet d'expressions faciales (une même personne de la base ORL).*

Cependant, en présence d'importantes variations de pose et d'éclairage les performances des systèmes se dégradent fortement, c'est pourquoi, ces deux problèmes en particulier sont considérés majeurs [96] et la littérature [60] reporte de nombreuses méthodes et techniques pour les corriger.

Selon Moses et al. [35]: " les variations dues à l'illumination et à la direction de la prise de vue, entre les images d'un même visage, sont plus importantes que celles entre les visages de personnes différentes ". Ces variations peuvent être dues à l'endroit d'acquisition des images (à l'intérieur ou à l'extérieur), à la direction et au nombre de sources de lumière utilisées,... Comme le montre la figure III.10, des variations d'illumination extrêmes peuvent engendrer une occultation d'une partie du visage. D'après l'évaluation du protocole FRVT 2000 [60], le passage d'un éclairage intérieur (ensemble d'apprentissage) à un autre extérieur (base de test) conduit à une baisse significative des performances (taux d'identification 0.55 pour le meilleur algorithme testé).

Pour corriger le problème des variations d'illumination, des techniques heuristiques, entre autres, ont été appliquées: la normalisation du contraste (contrast normalisation) [111],

la normalisation d'intensité et l'égalisation d'histogramme [90] et l'élimination des 3 premières composantes principales qui capturent les variations d'éclairage [49][57].

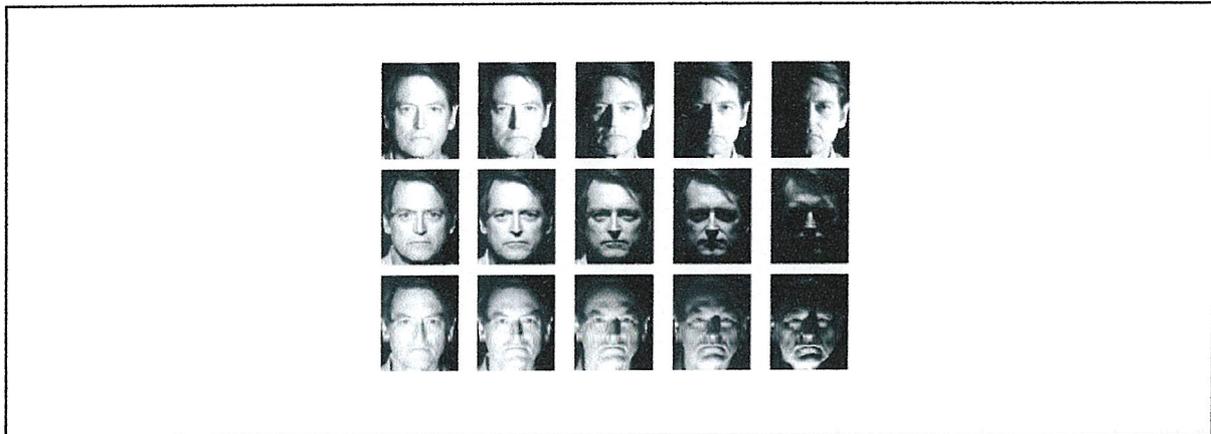


Figure III.10. Extrait de la base Harvard montrant l'effet de l'éclairage sur la même personne [57].

D'après l'évaluation du FRVT 2000, les variations de pose entre $\pm 25^\circ$ n'affectent pas significativement les performances des systèmes par contre à partir de $\pm 40^\circ$ (figure III.11), la dégradation des performances est significative. Pour traiter ce problème, on distingue entre les approches basées sur plusieurs vues (Multiview-based Approches) [92], les approches hybrides (Hybride Approches) [48][93] et les approches basées sur une seule image (Single-image based approaches) [95]. Ces dernières, ne motivent pas vraiment les chercheurs à cause de leur complexité et coût de calcul élevé [60].

Parmi les méthodes de la 1^{ère} catégorie celle de Beymer [92] qui consistait à représenter les visages par des modèles formés à partir de différentes vues de l'intervalle ($-30^\circ, -20^\circ, +20^\circ, +30^\circ$) et à comparer le visage inconnu avec uniquement les modèles de la même vue. Cette méthode arrivait à reconnaître 98% d'une base de 62 personnes avec 15 modèles pour l'apprentissage et 10 pour le test. Cependant les approches de cette catégorie [60]: **i)** nécessitent d'avoir une base de visages contenant plusieurs vues par personnes ce qui n'est pas toujours disponible, **ii)** ne traitent pas le problème des variations de l'éclairage et d'expression faciale et **iii)** sont coûteuses en temps de calcul.

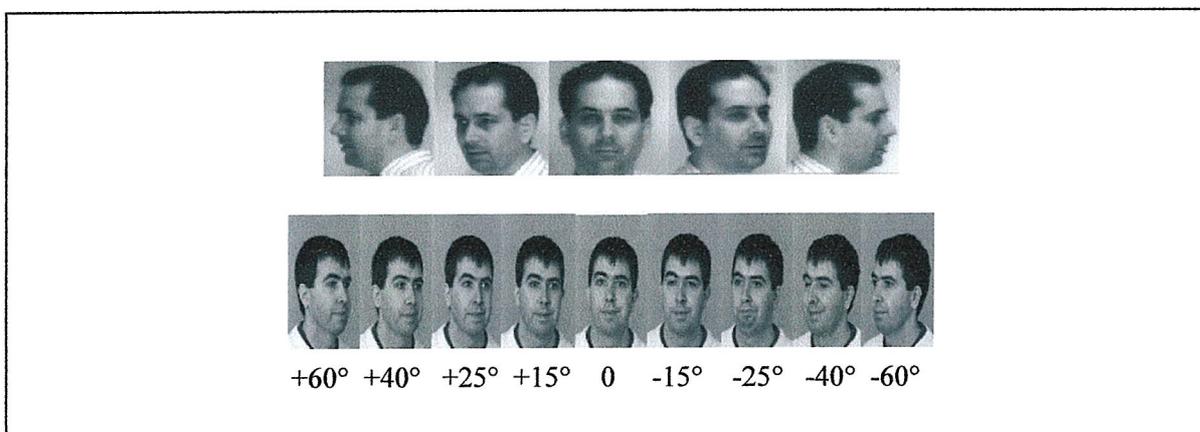


Figure III.11. 1^{ère} ligne: plusieurs vues d'un visage utilisées pour appliquer View-based eigenspace [48] et 2^{ème} ligne: un visage de la base FERET avec des vues orientées.

Quant aux approches hybrides, elles sont les plus populaires, et nécessitent plusieurs vues pour l'apprentissage tandis qu'une seule vue pour le test peut suffire. Parmi lesquelles, «View-Based Eigenspace » [48] ; une extension de la méthode eigenfaces. Elle consiste à représenter chaque vue par un sous espace capturant la variation des individus présentant la même vue ce qui donne M sous espaces pour M vues. Pour évaluer leur méthode, Pentland et al. ont effectué deux expériences sur une base de 21 personnes avec 9 vues pour chacun. Pour le 1^{er} test, les vues de l'intervalle ($\pm 90^\circ$, $\pm 45^\circ$, 0) ont été utilisées pour l'apprentissage et celles de ($\pm 68^\circ$, $\pm 23^\circ$) pour le test ce qui a abouti à un taux de 90%. La 2^{ème} expérience a donné 83% avec des vues appartenant à l'intervalle (-90° , 0, -45°) pour l'apprentissage et allant de $+68^\circ$ à $+90^\circ$ pour le test.

II.8 Extraction des éléments caractéristiques du visage

Les différents éléments faciaux (yeux, nez, bouche,...) sont d'une importance primordiale pour la détection et le suivi de visages, la reconnaissance des expressions faciales et la RdV. Toutes ces applications sont d'autant plus performantes que la localisation de ces éléments est plus précise. Souvent la localisation des éléments et la détection de visages s'effectuent simultanément c'est-à-dire détecter un visage en localisant ses éléments ou localiser les éléments pour détecter un visage.

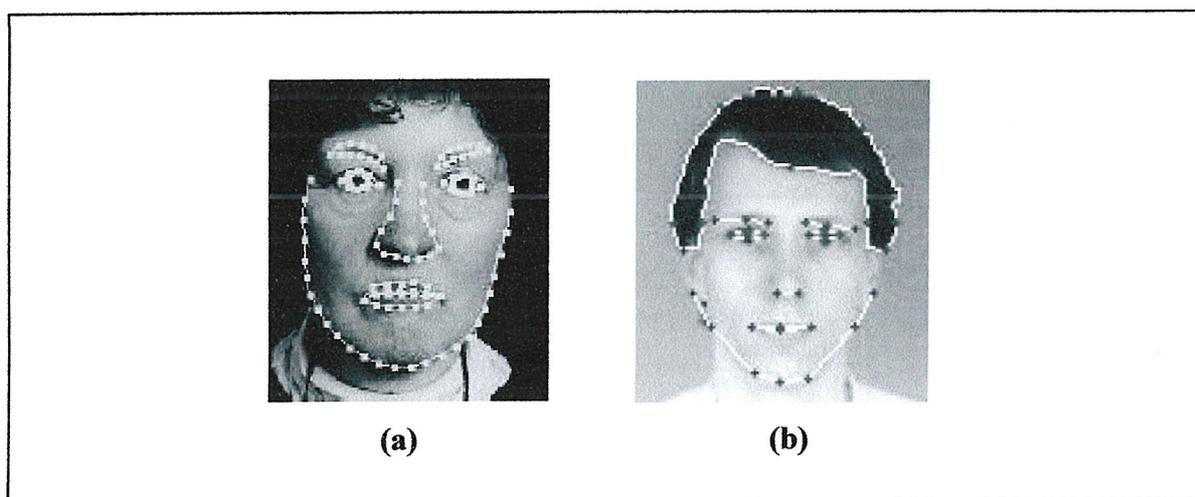


Figure III.12. Exemples de localisation de 122 (a) et 40 (b) points caractéristiques.

La localisation peut se faire manuellement [65] ou automatiquement [74][102], à partir des images de vue de face [46] (figure III.12) ou de visages orientés [48][23] (figure III.14). Dans tous les cas, un algorithme de localisation renvoie des points caractéristiques (coins des yeux et centre des iris, centre de la bouche, narines, commissures,...) [62][63][64] ou des contours [38][64][80][67]. Il est également possible d'extraire des régions contenant les différents éléments faciaux [48][81].

Les yeux sont les plus visés par ces algorithmes vu qu'ils sont discriminants [74], servent à aligner les images de visages avant de les utiliser [58] et donnent une estimation de l'orientation de la tête suivis de la bouche et du nez. Ces derniers dépendent souvent de la localisation des yeux si l'algorithme est basé sur les relations géométriques du visage ou sur l'intensité des images [46][63].

Généralement, on peut répertorier les approches de localisation des éléments faciaux en trois catégories [60]:

Méthodes basées sur les modèles de caractéristiques (Feature template-based methods) qui consistent à assigner des modèles prédéfinis ou déformables [73] aux différents attributs faciaux, parmi lesquelles les approches proposées par Hallinan [100] et Funabiki et al. [80] pour détecter les yeux et par Yuille et al. [101] pour les yeux et la bouche. Les limitations de cette catégorie sont dues principalement à la variance des éléments à modéliser (yeux fermés, bouche ouverte,...) et à leur occultation.

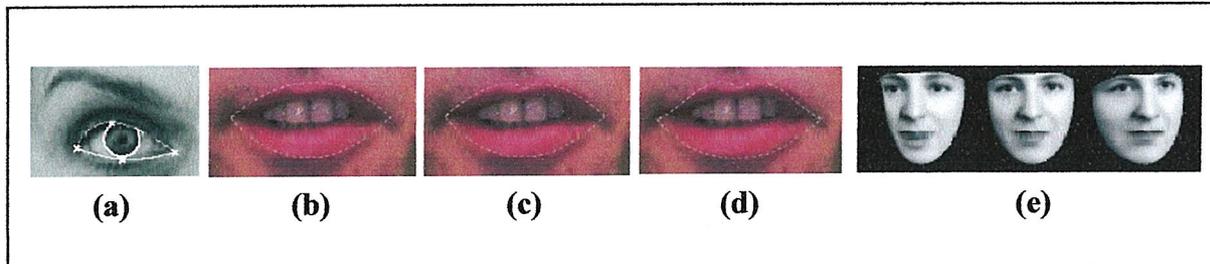


Figure III.13. (a) modèle à une courbe de Bézier pour le contour supérieur et une parabole pour le contour inférieur, (b) modèle à 2 paraboles, (c) modèle à 3 paraboles, (d) modèle à base de quartique [64] et (e) modèle actif d'apparence [98].

Méthodes basées sur la similitude structurelle (Structural matching): comparées aux précédentes, ces méthodes sont plus robustes pour la détection des éléments du visage. Parmi lesquelles on compte le modèle actif de forme (Active Shape Model) de Cootes et al.95 [97], le modèle flexible d'apparence (Flexible Appearance Model) proposé par Lanitis et al.95 [99] et le modèle actif d'apparence (Active Appearance Models) de Cootes et al. .98 [98].

Méthodes génériques (Generic Method): elles sont basées sur les lignes et les contours (edges) comme celle proposé par Craw et al.87 [78] et sur les contours actifs [46][67] dont l'inconvénient majeur est l'initialisation des courbes.

Dans la représentation eigenfeatures [48], l'erreur de reconstruction : DFFS (Distance From Face Space) a été utilisée pour détecter les yeux, le nez et la bouche à partir des vues de face. Son extension ; la représentation View-based eigenspace a permis également une détection à partir des visages orientés ($\pm 68^\circ$, $\pm 23^\circ$). Le taux de détection des yeux était 90%, cependant, il se dégrade dans les cas du nez et de la bouche à cause des variations dans la forme et la position dues à l'orientation.

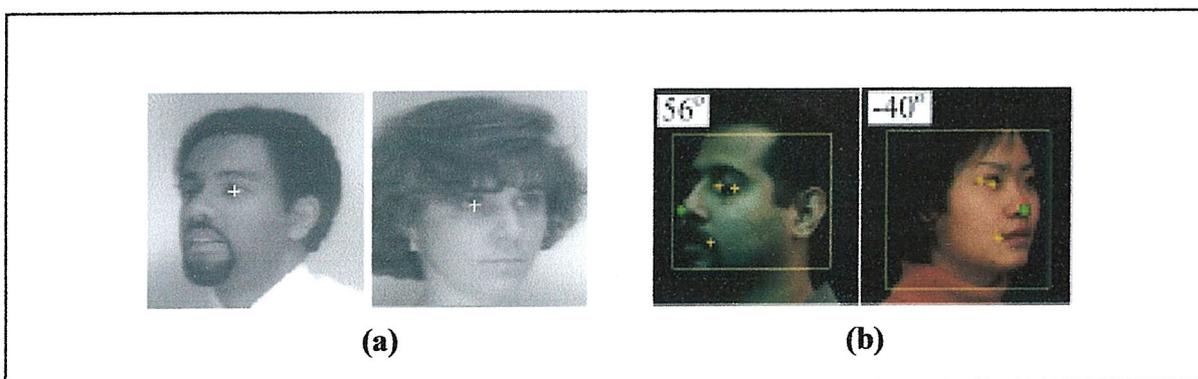


Figure III.14. (a) Détection des yeux par DFFS [48] et (b) Détection des éléments du visage à partir de visages orientés [23].

II.10 Evaluation des systèmes de reconnaissance automatique de visages

La littérature reporte de nombreux systèmes de reconnaissance de visages, chacun possède des avantages et des inconvénients et destiné à une application spécifique. L'avancement des recherches dans le domaine de la RdV, la disponibilité d'outils informatiques et de différentes bases de visages ont incité à l'évaluation de tous les systèmes réalisés. La première évaluation remonte à août 1994 avec le protocole FERET [96] qui a attiré neuf institutions et compagnies. Il visait le test des algorithmes destinés à localiser, normaliser et identifier automatiquement les visages dans les contextes de vérification et d'identification. La 2^{ème} et la 3^{ème} évaluation étaient réalisées en mars 1995, septembre 1996 et mars 1997 respectivement. Les tests menés sur la base FERET ont démontré la supériorité de trois algorithmes à savoir : probabilistic cigenfaces de MIT, souspace LDA [58][59] de UMD et Elastic Graph Matching [93] de USC.

Les protocoles FRVT 2000 et FRVT 2002 (Face Recognition Vendor Test) sont deux autres évaluations de logiciels mis au point par des laboratoires et des industries : cinq compagnies (Banque-Tec International Pty.Ltd, C-Vis Computer Vision and Automation GmbH, Miros Inc., Lau Technologies and Vision Corportaion) ont participé au premier, et dix participants (Identix, Viisage Technology, Imagis Technologies Inc., Iconquest, ...) au deuxième. Pour cela la base de visages HCInt comportant 121589 images de 37437 personnes a été utilisée pour tester les performances des systèmes soit en mode d'identification, de vérification ou en celui de watch list.

En plus des variations de l'éclairage, de la pose de la tête, le protocole FRVT 2002, reporte d'autres facteurs influant sur la performance des systèmes de reconnaissance faciale à savoir le temps, l'âge, le sexe et la taille de la base de visages. Un an d'écart entre les images d'apprentissage et celles de test, entraîne une baisse de 5%. Les taux de reconnaissance des hommes dépassent ceux des femmes. Pour l'âge, les meilleurs systèmes ont donné un taux de 74% (en moyenne) pour les personnes âgés de (38-42) ans, et 62% pour les jeunes entre (18-22) ans. Quant à la taille, les taux sont 85% pour 800 personnes, 83% pour 1600 et 73% pour 37437 personnes.

En résumé, l'évaluation d'un système de reconnaissance de visages peut dépendre:

1. **de la méthode appliquée** (globale, basée sur les éléments du visage ou hybride): les méthodes basées sur les éléments du visage nécessitent des images de grandes tailles et des algorithmes précis de détection et de localisation, tandis que les méthodes globales exigent des méthodes d'extraction de caractéristiques et de réduction de dimensions (voir l'annexe D),
2. **de la base de visages utilisée**: les résultats diffèrent, pour une même méthode, selon la taille et les conditions d'acquisition des images de visages (éclairage intérieur ou extérieur, expressions faciales, occultations,...) (voir l'annexe D),
3. **du nombre d'images pour chaque personne**: l'ADL est plus performante que l'ACP, mais ne peut pas s'appliquer si la base contient une seule image par personne. La performance du classificateur dépend également du nombre d'images par personne utilisées pour l'apprentissage,
4. **du prétraitement et de la mesure de similarité** (dans le cas de la règle du plus proche voisin): l'ADL appliquée sur des images centrées réduites (ZMUV : Zero Mean

Unit Variance) est plus performante que l'ADL avec une égalisation d'histogramme [58][59]. L'ACI appliquée avec la distance cosinus (cosine distance) au lieu de la distance euclidienne est plus performante que l'ACP [72][84]. Cependant avec la norme L_2 , les deux méthodes sont similaires, ceci peut être justifié, en plus de la distance utilisée, par l'élimination des arrières plans et des cheveux [53][61].

Conclusion

Parmi les diverses techniques biométriques existantes, nous avons choisi une caractéristique qualifiée de naturelle et amicale : c'est le visage. Ainsi, nous avons mené, dans ce chapitre, une étude bibliographique qui nous a permis de passer en revue différentes approches proposées, anciennes et récentes et de se rendre compte de la richesse du domaine de la reconnaissance automatique de visages depuis les premiers travaux du japonais Kanade qui date de 1973 jusqu'aujourd'hui.

En résumé, nous pouvons conclure que malgré l'avancement des recherches, reconnaître automatiquement une personne, à partir de son visage, reste une tâche de reconnaissance des formes difficile compte tenu les variations inter et intra visages dues, respectivement, à la forme du visage et de ses composantes, aux expressions faciales et au vieillissement par exemple.

Le choix d'une méthode pour identifier ou authentifier une personne dépend, d'une part, de la qualité, de la taille et des conditions d'acquisition de la base de visages disponible, et d'autre part, du choix entre une approche globale ou basée sur les éléments faciaux : appliquer la 1^{ère} exige la correction du problème de la malédiction de dimensions par une des méthodes d'extraction de caractéristiques et de réduction de dimensions, et appliquer la 2^{ème} nécessite la disposition d'un algorithme précis pour localiser les différents éléments faciaux. Enfin, notre contribution fera l'objet du chapitre suivant,...

Chapitre IV

Approches Proposées & Résultats

Introduction

En reconnaissance de visages, l'application des méthodes classiques d'extraction de caractéristiques telles que l'ACP et l'ADL exige le passage par la concaténation des lignes ou des colonnes des images de visages pour obtenir des vecteurs. Cette transformation donne des vecteurs de très grandes dimensions : pour la base ORL, par exemple, les images originales sont de tailles (112×92) , leur transformation donne ainsi des vecteurs de tailles (10304×1) . Le traitement de ces vecteurs sera une source de problèmes.

Notre contribution porte sur deux parties, la première consiste à proposer une approche de compression et de codage de données que nous l'avons nommée **Neuro-ACPDL2D** basée sur une projection hybride des matrices-visages pour l'extraction de caractéristiques et la classification de celles-ci par un PMC. Cette approche offre une compression qui dépasse celle des autres méthodes bidimensionnelles tout en garantissant un taux de reconnaissance satisfaisant.

Il nous arrive de reconnaître une personne à partir, seulement, de ses caractéristiques faciales marquées : de beaux yeux, un long nez, de grandes oreilles, une grande bouche, etc. En raison du pouvoir discriminant des éléments faciaux et la performance des méthodes globales bidimensionnelles, la deuxième partie de notre contribution vise à améliorer l'identification en présence d'occultations partielles en proposant les **approches locales bidimensionnelles** dont le principe est de combiner les approches globales et les éléments du visage. Cela en appliquant séparément l'ACP2D, l'ADL2D et l'approche hybride ACPDL2D sur les régions des yeux (avec les sourcils) et du nez au lieu du visage entier.

La comparaison, sur la même base de visages (B-LAIG), de la méthode géométrique [69] et des méthodes proposées dans ce travail nous a permis de tirer d'importantes conclusions,...

I. Approche Hybride Bidimensionnelle Neuro-ACPDL2D et Approches Locales bidimensionnelles

I.1 Méthodes bidimensionnelles d'extraction de caractéristiques

Parmi les approches de reconnaissance de visages qui ont attiré le plus notre attention; l'analyse en composantes principales bidimensionnelle (ACP2D) [52] et l'analyse discriminante linéaire bidimensionnelle (ADL2D) [54][87]. Ces dernières reposent sur une projection linéaire dans un espace bidimensionnel pour la phase d'extraction de caractéristiques (signatures) et la règle du plus proche voisin (Nearest Neighbor : NN) pour la classification. Avant de les détailler, nous schématisons leur principe général (figure IV.1) :

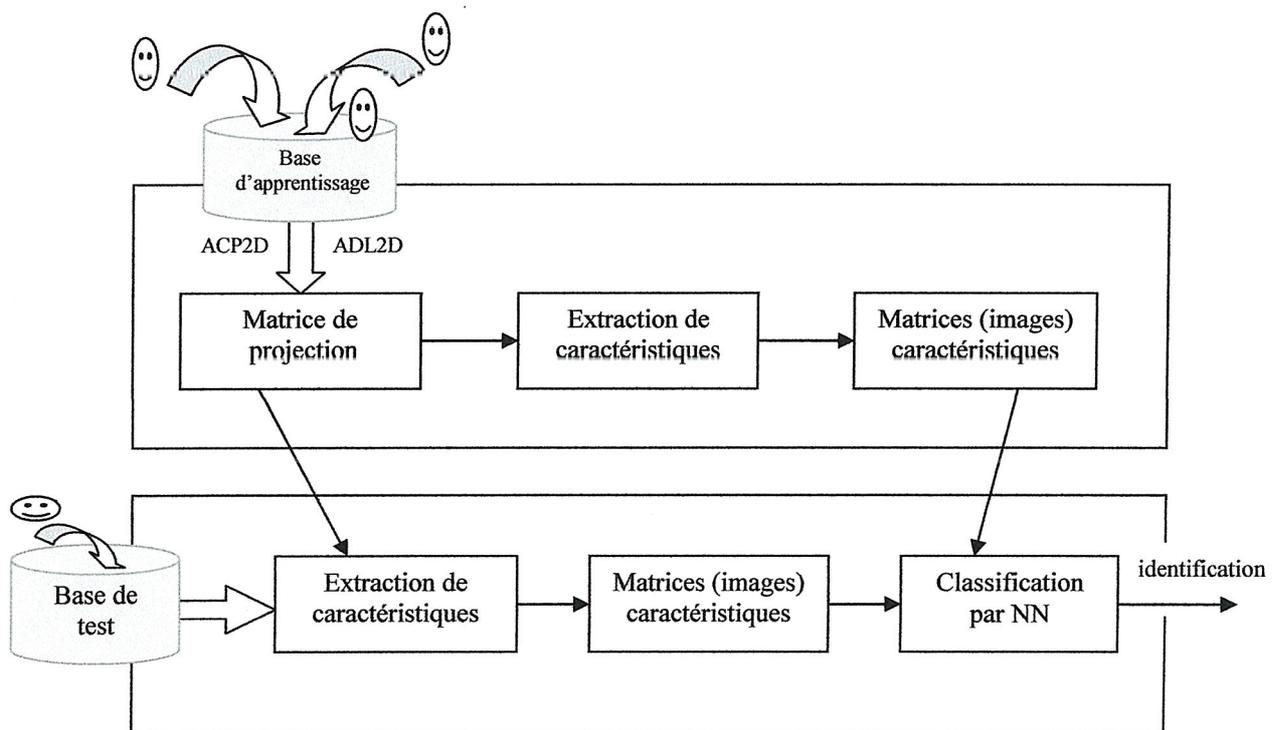


Figure IV.1. Principe de l'ACP2D et l'ADL2D.

I.1.1 Analyse en composantes principales bidimensionnelle (ACP2D)

Proposée par J.Yang en 2004 [52], l'ACP2D est la première méthode d'extraction de caractéristiques et de réduction de dimensions basée sur l'ACP mais qui traite directement les images de visages comme des matrices sans avoir à les transformer en vecteurs comme l'exige les anciennes méthodes globales. Cette méthode a été testée sur trois bases de visages internationales :

- ORL pour évaluer ses performances en présence des variations dans les conditions d'illumination, dans la pose de la tête, dans l'expression faciale et dans l'échelle du visage,
- la base AR présentant des variations dans le temps, dans l'expression faciale et dans les conditions d'éclairage et
- la base de Yale acquise avec des conditions d'illumination et d'expressions faciales variables.

Les auteurs reportent que l'ACP2D présente une performance élevée en termes **i)** de reconnaissance (voir l'annexe D) comparée à eigenfaces [51][50], Fisherfaces [57], ICA et kernel PCA [76], **ii)** de temps de calcul ; plus le nombre d'images par personne est grand, plus la différence est très importante (voir tableau IV.1 et tableau IV.2) et **iii)** de reconstruction des images originales, comparée à eigenfaces. Cependant la supériorité de l'ACP2D n'a pas été théoriquement justifiée dans leur article.

Images/classe	1	2	3	4	5
eigenfaces	44.45	89.00	139.36	198.95	304.61
ACP2D	10.76	11.23	12.59	13.40	14.03

Tableau IV.1 Comparaison entre les temps de calcul (seconde) de la phase d'extraction de caractéristiques obtenus pour la base ORL [52].

	Variations dans le temps	Variations des expressions faciales	Variations de l'éclairage (moyenne)
eigenfaces	434.87	130.42	129.56
ACP2D	16.26	7.25	8.32

Tableau IV.2 Comparaison entre les temps de calcul (seconde) de la phase d'extraction de caractéristiques obtenus pour la base AR [52].

I.1.1.1 Principe

Pour un ensemble d'apprentissage de M matrices de visages centrées, l'idée de l'ACP2D [52] est de projeter une matrice $X(n \times m)$ via une transformation linéaire telle que:

$$Y_i = X \cdot R_i \quad (\text{IV.1})$$

Où $Y_i (n \times 1)$ est dit vecteur composante principale et $R_i (m \times 1)$ est le vecteur de projection.

Le vecteur optimal R_i de la projection est obtenu en maximisant le critère de la variance totale généralisé suivant:

$$J(R) = R^T \cdot G_i \cdot R \quad (\text{IV.2})$$

Où $G_i (m \times m)$ est la matrice de covariance des images donnée par :

$$G_i = \frac{1}{M} \sum_{j=1}^M (X_j - \bar{X})^T (X_j - \bar{X}) \quad (\text{IV.3})$$

avec X_j : la $j^{\text{ème}}$ image de la base d'apprentissage.

\bar{X} : l'image moyenne totale de toutes les images de la base d'apprentissage.

En général un seul axe de projection optimal n'est pas suffisant. On doit sélectionner un ensemble d'axes de projection tel que:

$$\begin{aligned} \{R_1, R_2, \dots, R_{d_1}\} &= \arg \max J(R) \\ R_i^T R_j &= 0, i \neq j, i, j = 1, \dots, d_1 \end{aligned} \quad (\text{IV.4})$$

Ces axes sont les vecteurs propres de la matrice de covariance G_i correspondant aux « d_1 » plus grandes valeurs propres. L'extraction des caractéristiques d'une image X par l'ACP2D se fait donc selon :

$$Y = X.R \quad (\text{IV.5})$$

où $R = [R_1 R_2 \dots R_{d_1}]$ est la matrice de projection et $Y = [Y_1 Y_2 \dots Y_{d_1}]$ est la matrice (ou l'image) caractéristique de l'image X .

I.1.1.2 Reconstruction des images originales

De façon similaire à la méthode des eigenfaces, l'ACP2D utilise, pour la reconstruction d'une image de visage, la matrice caractéristique et les vecteurs propres de la matrice G_i à partir de l'équation (IV.5) comme suit:

$$\tilde{X} = Y.R^T = \sum_{k=1}^{d_1} Y_k.R_k^T \quad (\text{IV.6})$$

La figure IV.2 illustre la comparaison entre la reconstruction d'une image de la base B-LAIG par trois différentes méthodes.

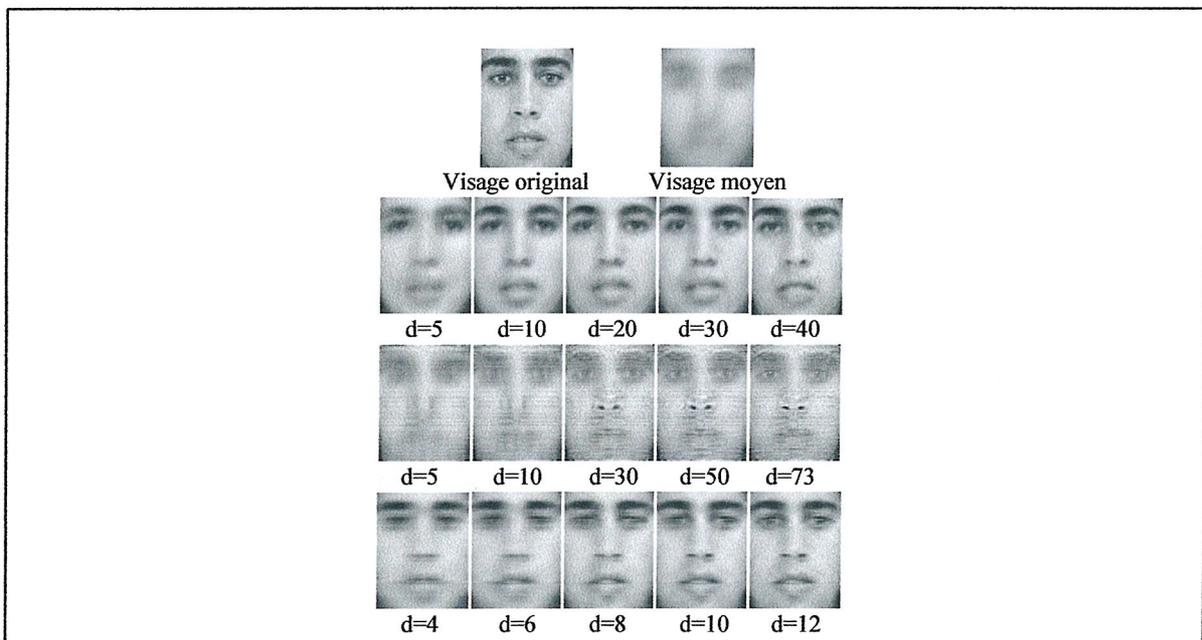


Figure IV.2. Comparaison entre les images reconstruites par : eigenfaces (1^{ère} ligne), ADL2DoC (2^{ème} ligne) et ACP2D (3^{ème} ligne).

Pour $d_1 = m$, l'ACP2D nous permet d'obtenir l'image originale tandis que pour $d_1 < m$, la reconstruction est approximative mais meilleure que celle par eigenfaces. Nous

avons remarqué également qu'elle est meilleure que la reconstruction par l'ADL2DoC dont le principe sera expliqué dans le paragraphe suivant. La reconstruction dans ce cas est avec perte même pour un nombre de vecteurs propres égale au nombre total de lignes de la matrice originale ($d_2 = n$).

I.1.1.3 Classification

Après l'obtention des matrices caractéristiques par l'ACP2D, la règle du plus proche voisin est utilisée pour leur classification. Pour ce faire, J.Yang [52] a proposé le calcul de distances entre deux matrices caractéristiques $Y_i = [y_1^{(i)} y_2^{(i)} \dots y_{d_1}^{(i)}]$ et $Y_j = [y_1^{(j)} y_2^{(j)} \dots y_{d_1}^{(j)}]$ comme suit :

$$d(Y_i, Y_j) = \sum_{k=1}^{d_1} \|y_k^{(i)} - y_k^{(j)}\|_2 \quad (\text{IV.7})$$

Où $\|y_k^{(i)} - y_k^{(j)}\|_2$ désigne la distance euclidienne entre les deux vecteurs composantes principales $y_k^{(i)}$ et $y_k^{(j)}$.

Pour un ensemble d'apprentissage de M matrices de visages où chacune est assignée à une identité (classe), une image test Y est affectée à la classe C_i de la matrice Y_i si:

$$d(Y, Y_j) = \min_j d(Y, Y_j) \quad (\text{IV.8})$$

Cette distance a été, également, adoptée par Visani et al. [54] et Kong et al. [83][87][91].

I.1.2 Analyse discriminante linéaire bidimensionnelle (ADL2D)

En s'inspirant de l'ACP2D, une autre méthode reposant sur l'ADL a été introduite avec différents noms à savoir : l'ADL2D orientée [54] qui se présente en deux versions; l'ADL2DoC (orientée en colonnes) et l'ADL2DoL (orientée en lignes) et l'ADF2D (Analyse Discriminante de Fisher Bidimensionnelle) avec ses deux algorithmes [91] : LU2DFDA et RU2DFDA (Left-multiplying Unilateral et Right-multiplying Unilateral 2D Fisher Discriminant Analysis respectivement). Cette méthode est plus performante que l'ADL et corrige le problème de la singularité de la matrice de covariance intra-classes ce qui la rend applicable pour les problèmes de petites tailles (Small Sample Size problem SSS) [87].

Pour une base d'apprentissage de M matrices de visages réparties en C classes où chaque classe contient n_c images de visages de tailles $(n \times m)$, le principe [54][91] des deux versions de l'ADL2D, résumé dans les paragraphes suivants, consiste à projeter toutes les images de la base sur une matrice de projection en maximisant un certain critère.

I.1.2.1 Analyse discriminante linéaire bidimensionnelle orientée en lignes (ADL2DoL)

L'ADL2DoL cherche à maximiser les variances inter-classes et minimiser les variances intra-classes en appliquant une projection linéaire à droite de la matrice de visage $X(n \times m)$ sur une matrice $P(m \times d_2)$ telle que:

$$W = X.P \quad (\text{IV.9})$$

Où la matrice $W(n \times d_2)$ est dite matrice caractéristique (ou matrice de signature associée à l'image X). La matrice de projection P est obtenue en maximisant le critère de Fisher généralisé suivant:

$$P^* = \arg \max_{P \in \mathfrak{R}^{(n \times d_2)}} \frac{|P^T S_b P|}{|P^T S_w P|} \quad (\text{IV.10})$$

Où S_b est la matrice de covariance inter-classes et S_w est la matrice de covariance intra-classes données respectivement par:

$$S_b = \sum_{c=1}^C n_c (\bar{X}_c - \bar{X})^T (\bar{X}_c - \bar{X}) \quad (\text{IV.11})$$

$$S_w = \sum_{c=1}^C \sum_{i=1}^{n_c} (X_i - \bar{X}_c)^T (X_i - \bar{X}_c) \quad (\text{IV.12})$$

avec \bar{X}_c : matrice moyenne de la classe c ,

\bar{X} : matrice moyenne totale (de toutes les matrices de la base d'apprentissage).

Les colonnes de la matrice P sont les vecteurs propres de la matrice $(S_w^{-1} S_b)$ associées aux « d_2 » plus grandes valeurs propres.

I.1.2.2 Analyse discriminante linéaire bidimensionnelle orientée en colonnes (ADL2DoC)

Dans ce cas, la projection de chaque matrice de visage X est effectuée à gauche telle que $Z = L^T . X$ est la matrice caractéristique de dimension $(d_2 \times m)$ où $L(n \times d_2)$ est la matrice de projection optimale maximisant le critère de Fisher généralisé :

$$L^* = \arg \max_{L \in \mathfrak{R}^{(n \times d_2)}} \frac{|L^T \Sigma_b L|}{|L^T \Sigma_w L|} \quad (\text{IV.13})$$

ayant pour colonnes, les d_2 premiers vecteurs propres de la matrice $(\Sigma_w^{-1} \Sigma_b)$ associés aux « d_2 » plus grandes valeurs propres. Σ_w et Σ_b désignent respectivement les matrices intra et inter classes généralisées de la base d'apprentissage:

$$\Sigma_w = \sum_{c=1}^C \sum_{i=1}^{n_c} (X_i - \bar{X}_c)(X_i - \bar{X}_c)^T \quad (\text{IV.14})$$

$$\Sigma_b = \sum_{c=1}^C n_c (\bar{X}_c - \bar{X})(\bar{X}_c - \bar{X})^T \quad (\text{IV.15})$$

I.2 Supériorité de l'ACP2D et de l'ADL2D orientée

Il a été démontré [83] que l'ACP2D est l'ACP appliquée sur les lignes de toutes les images de visages et que LU2DFDA et RU2DFDA [87][91] sont l'ADL appliquée respectivement sur les colonnes et les lignes des matrices de visages (voir l'annexe C). Par conséquent, la supériorité de l'ACP2D sur l'ACP et celle de l'ADL2D sur l'ADL peut être justifiée par les arguments suivants:

- Calcul directe et facile de la matrice de covariance de dimension $(m \times m)$ ou $(n \times n)$ au lieu de $(n \times m, n \times m)$ ou même $(M \times M)$ comme l'a proposé Turk et Pentland [50] pour réduire le temps et la complexité de calcul des eigenfaces,
- Préservation de l'information spatiale bidimensionnelle [83],
- Utilisation, pour la classification, de la somme des distances euclidiennes entre les vecteurs composantes principales formant la matrice caractéristique d'une image test et leur similaires des images d'apprentissage.

1.3 Inconvénients de l'ACP2D et de l'ADL2D orientée

Malgré les avantages offerts par ces deux méthodes, elles souffrent de quelques inconvénients à savoir :

- Elles nécessitent plus de coefficients pour représenter les visages à reconnaître que l'ACP (eigenfaces) et l'ADL,
- Elles effectuent une projection unilatérale (multiplication à droite ou à gauche) réduisant les redondances au niveau des lignes (ou des colonnes) seulement [83][91],
- Elles dépendent du nombre de vecteurs propres à retenir qui lui-même dépend de la base de visages utilisée.

L'application de ces deux méthodes : l'ACP2D et l'ADL2DoC sur la base B-LAIG, nous a permis d'obtenir les courbes de la figure IV.3 qui illustrent la décroissance des valeurs propres.

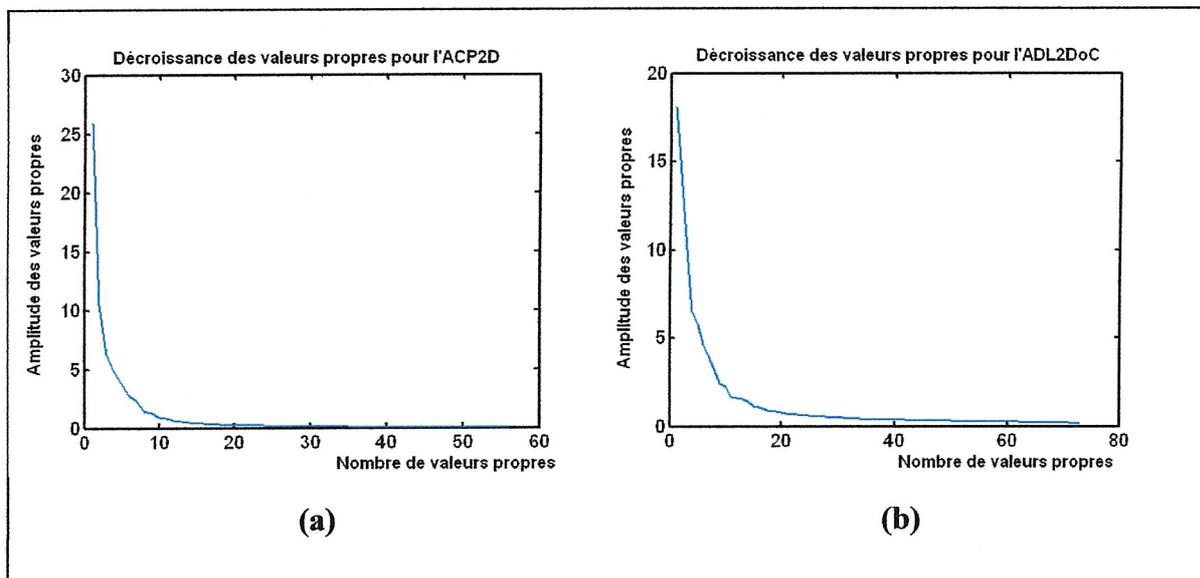


Figure IV.3. Décroissance des valeurs propres obtenues par : l'ACP2D (a) et l'ADL2DoC (b).

L'énergie d'une image est concentrée dans les premiers vecteurs propres, c'est pourquoi ces derniers sont utilisés pour la reconnaissance. Les approches bidimensionnelles restent souffrantes du problème du nombre optimal des vecteurs propres à retenir. Sur la base ORL, l'ACP2D est optimale pour 3 vecteurs propres [54] et l'ACP2D bilatérale pour 5 [91].

I.4 L'approche Hybride Bidimensionnelle proposée : Neuro-ACPD2D

I.4.1 Idée et motivations

Le principe de notre système de reconnaissance automatique de visages (figure IV.4) qui comporte les trois étapes suivantes:

- localisation automatique de visages,
- extraction de caractéristiques par l'approche hybride bidimensionnelle,
- reconnaissance neuronale de visages,

consiste à compacter le nombre de coefficients nécessaires pour représenter une matrice de visage X en combinant l'ACP2D et l'ADL2DoC et effectuer la classification par un PMC.

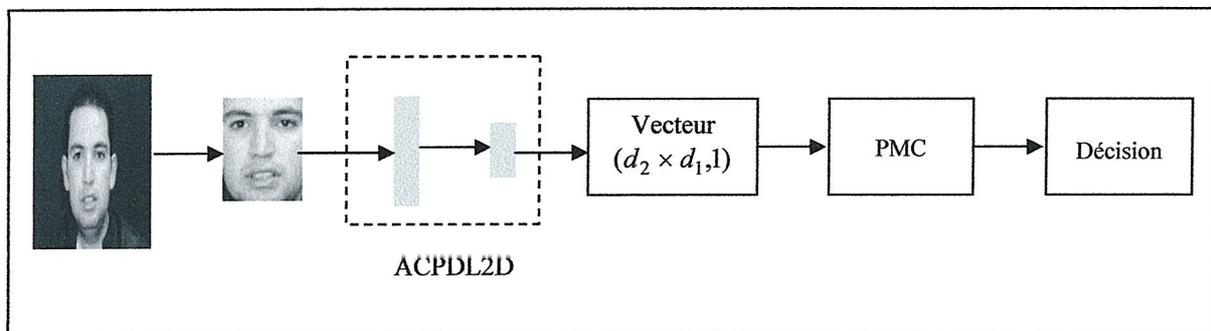


Figure IV.4. Système Neuro-ACPD2D.

Via la fusion proposée, nous bénéficions à la fois du pouvoir représentatif de l'ACP2D, du pouvoir discriminant de l'ADL2D et du pouvoir de généralisation du PMC dans l'objectif d'arriver à un système amélioré corrigeant deux inconvénients de ces deux méthodes bidimensionnelles. Ainsi, l'approche proposée réduit le nombre important de coefficients nécessaires à la représentation des images de visages en effectuant une projection hybride bilatérale pour éliminer les redondances au niveau des lignes et des colonnes au même temps ce qui justifie le choix de fusionner l'ACP2D avec l'ADL2DoC et non pas avec l'ADL2DoL.

A notre connaissance, l'ACP2D et l'ADL2DoC étaient, jusqu'à présent, testées avec la règle du plus proche voisin et nous avons fait l'initiative de les appliquer avec un PMC avant de les combiner. Les étapes de notre système seront détaillées dans les paragraphes qui suivent :

I.4.2 Localisation automatique de visages

Au contraire de certains systèmes de reconnaissance faciale qui utilisent les images originales avec les arrières plans et les cheveux ou segmentent manuellement les visages [49][57][58][52][77], nous proposons de précéder l'extraction de caractéristiques d'une étape de localisation automatique pour éliminer les arrières plans, les voiles et les cheveux qui peuvent influencer la reconnaissance. En se basant sur la méthode de [81] pour détecter les régions des yeux, nous avons réalisé cette première étape qui rend l'identification basée uniquement sur la forme du visage segmenté.

La figure IV.5 schématise les différentes étapes de cette localisation, elle retourne de nouvelles images qui ne contiennent que les visages et sont de dimensions (73×56) .

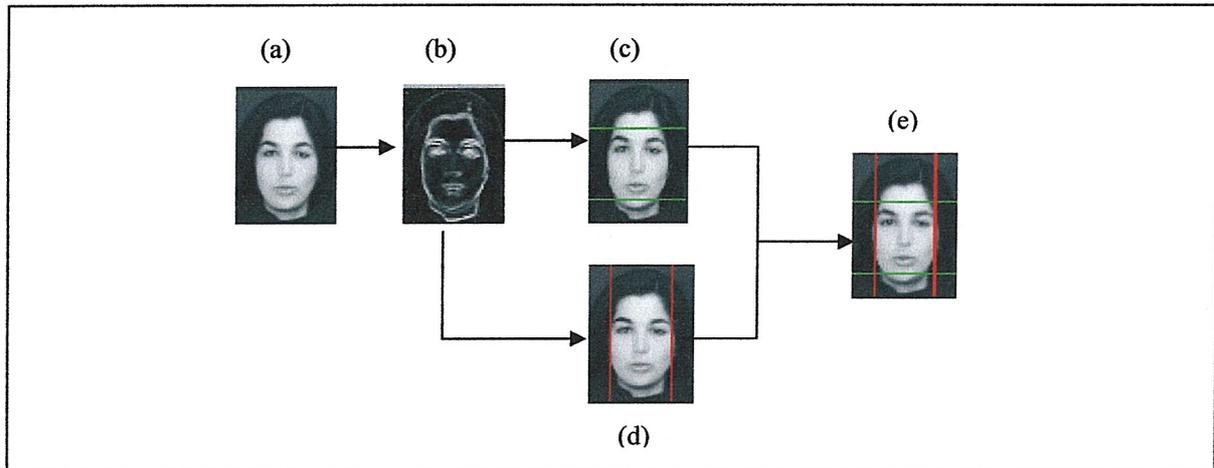


Figure IV.5. Localisation automatique de visages.

- (a) Image originale
- (b) Norme du gradient de l'image
- (c) Limites horizontales obtenues par une projection horizontale
- (d) Limites verticales obtenues par une projection verticale

Les projections verticales et horizontales du gradient de l'image originale, de taille $(n \times m)$, vérifient respectivement :

$$H_v(i) = \sum_{j=1}^n I_g(i, j) \quad \text{pour } 1 \leq i \leq m \quad (\text{IV.16.a})$$

$$H_h(j) = \sum_{i=1}^m I_g(i, j) \quad \text{pour } 1 \leq j \leq n \quad (\text{IV.16.b})$$

Les limites horizontales représentées en figure IV.5(c) sont données par :

$$y_H = \arg \max_j (H_h(j)) + l_1 \quad \text{avec } 1 \leq j \leq n/4 \quad (\text{IV.17.a})$$

$$y_L = \arg \max_j (H_h(j)) + l_2 \quad \text{avec } 3.n/4 \leq j \leq n \quad (\text{IV.17.b})$$

Où l_1 et l_2 sont deux constantes déterminées expérimentalement pour assurer que la limite supérieure y_H est au dessus des sourcils et que la limite inférieure y_L est au dessous de la lèvre inférieure.

Quant aux limites verticales (figure IV.5(d)), elles sont données par :

$$x_L = \arg \max_i (H_v(i)) \quad \text{avec } 1 \leq i \leq m/3 \quad (\text{IV.18.a})$$

$$x_R = \arg \max_i (H_v(i)) \quad \text{avec } 2.m/3 \leq i \leq m \quad (\text{IV.18.b})$$

I.4.2 Extraction de caractéristiques et réduction de dimensions par l'ACPDL2D

L'extraction de caractéristiques (signatures) est la deuxième étape du système proposé. A cet objectif, nous proposons une approche qui s'inscrit dans les approches globales bidimensionnelles. Elle vise la réduction de dimensions et le codage de données originales.

Son principe consiste à effectuer, pour chaque image de visage $X(n \times m)$ de l'ensemble d'apprentissage, une projection bilatérale hybride en appliquant premièrement, une projection à droite à l'aide de l'ACP2D suivie d'une deuxième projection à gauche en appliquant l'ADL2DoC, autrement dit nous appliquons l'ADL2DoC sur les matrices caractéristiques obtenues par l'ACP2D. Cette nouvelle représentation, que nous avons noté **ACPDL2D** (Analyse en Composantes Principales Discriminante Linéaire bidimensionnelle), donne de nouvelles matrices de tailles $(d_2 \times d_1)$, dites images caractéristiques, au lieu de $(n \times d_1)$ pour l'ACP2D ou $(d_2 \times m)$ pour l'ADL2DoC appliquées séparément.

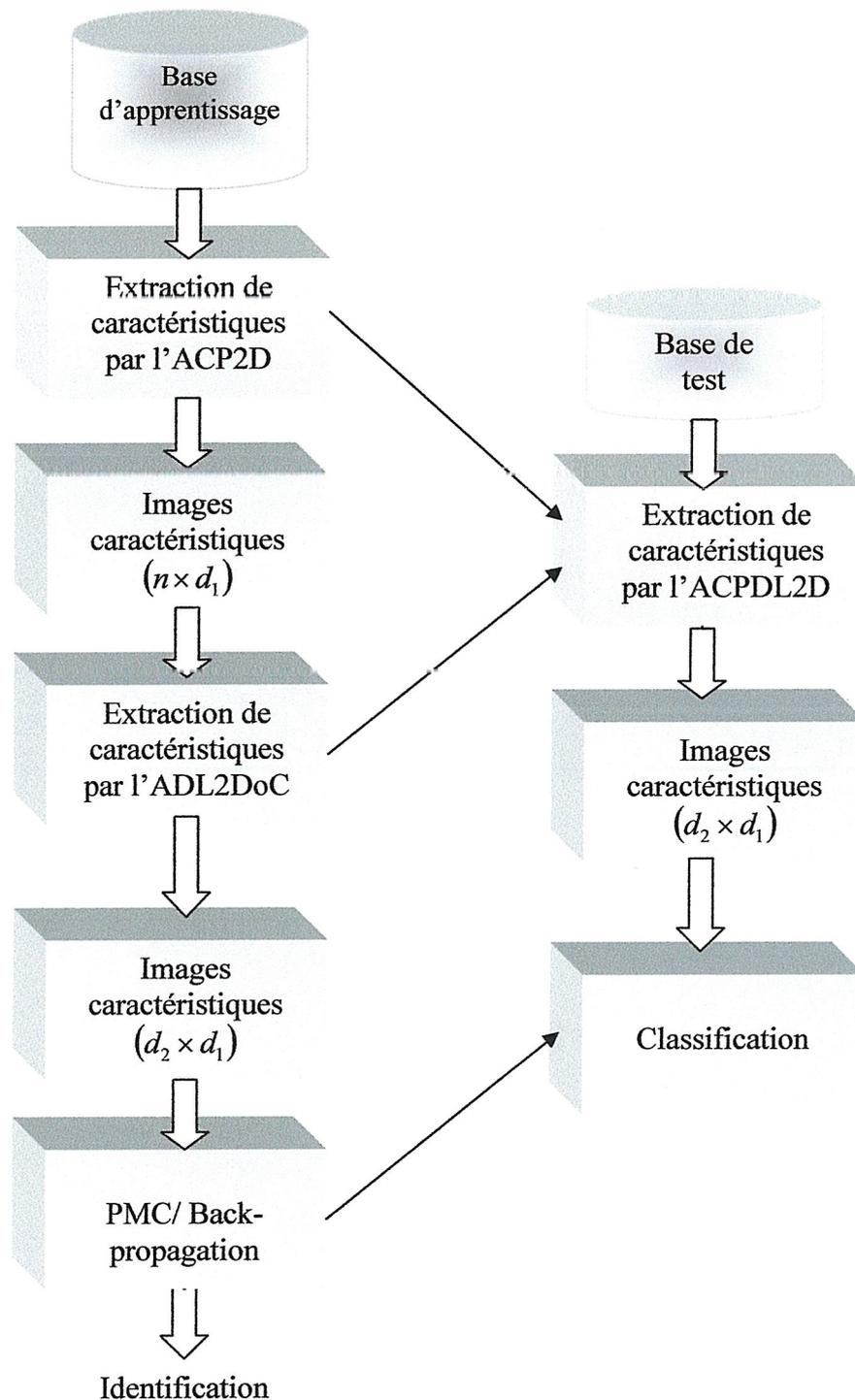
Les paramètres d_1 et d_2 sont les nombres de vecteurs propres retenus respectivement pour l'ACP2D et l'ADL2DoC. L'organigramme IV.1 et l'algorithme IV.1 (page suivante) détaillent les étapes de l'approche hybride.

I.4.3 Classification de visages

Une fois les matrices de projection et les matrices caractéristiques sont obtenues par l'application de l'approche ACPDL2D, la reconnaissance de visages devient une tâche de reconnaissance de formes. Les matrices caractéristiques doivent être transformées en vecteurs avant de les fournir au PMC dont l'algorithme d'apprentissage est celui du backpropagation détaillé par l'algorithme II.1 (voir chapitre II).

La phase d'apprentissage revient alors à trouver le nombre de neurones dans la couche cachée et les paramètres optimaux du PMC. La couche de sortie du PMC comporte 40 neurones (nombre de classes de la base de visages) et celle d'entrée contient $(n \times d_1)$ neurones dans le cas de l'ACP2D, $(d_2 \times m)$ neurones pour l'ADL2DoC et $(d_2 \times d_1)$ neurones pour l'ACPDL2D. Quant à la phase de test, elle consiste à projeter les images de visages à reconnaître dans le nouveau sous espace réduit via les matrices de projection déterminées dans la phase d'apprentissage et effectuer ensuite la classification à l'aide du PMC.

I.4.4 Organigramme et algorithme de l'approche Neuro-ACPDL2D



Organigramme IV.1. Principe de l'Approche Neuro-ACPDL2D.

Cet algorithme détaille les différentes étapes de l'approche proposée Neuro-ACPD2D

Algorithme IV.1: Neuro-ACPD2D

Entrées : M images de visages

Sorties : matrices de projection et matrices caractéristiques

1. Former un ensemble de M images de visages: $P = [X_1 X_2 \cdots X_M]$
 2. Calculer la matrice moyenne de toutes les images: $\bar{X} = \frac{1}{M} \sum_{j=1}^M X_j$
 3. Calculer la matrice de covariance: $G_t = \frac{1}{M} \sum_{j=1}^M (X_j - \bar{X})^T (X_j - \bar{X})$
 4. Calculer les valeurs propres λ_k et les vecteurs propres R_k de la matrice $G_t (m \times m)$
 5. Retenir d_1 vecteurs propres triés en ordre décroissant des valeurs propres, Rassembler ces d_1 vecteurs propres dans une matrice: $R = [R_1 R_2 \cdots R_{d_1}]$
 6. Projeter les matrices de visages centrées dans un nouvel espace de visages: $Y_j = \hat{X}_j \cdot R$ pour $j = 1, \dots, M$
 7. Reconstruire les images de visages: $X_j = Y_j \cdot R^T + \bar{X}$ (pour vérification)
 8. Former une nouvelle base d'apprentissage Ω formée des matrices caractéristiques $Y_j (n \times d_1)$,
 9. Calculer la moyenne de chaque classe de la base Ω : $\bar{Y}_c = \frac{1}{n_c} \sum_{i=1}^{n_c} Y_i$ pour $c = 1, \dots, C$
 10. Calculer la moyenne totale de la base Ω : $\bar{Y} = \frac{1}{M} \sum_{j=1}^M Y_j$,
 11. Calculer les matrices de covariance intra et inter classes:

$$\Sigma_w = \sum_{c=1}^C \sum_{i=1}^{n_c} (Y_i - \bar{Y}_c)(Y_i - \bar{Y}_c)^T \quad \text{et}$$

$$\Sigma_b = \sum_{c=1}^C n_c (\bar{Y}_c - \bar{Y})(\bar{Y}_c - \bar{Y})^T$$
 12. Calculer les valeurs et les vecteurs propres de la matrice $(\Sigma_w^{-1} \cdot \Sigma_b)$ de taille $(n \times n)$,
 13. Trier ces valeurs propres en ordre décroissant et retenir d_2 vecteurs propres correspondant à d_2 valeurs propres les plus grandes, Former une matrice dont les colonnes sont les d_2 vecteurs propres: $L = [L_1 L_2 \cdots L_{d_2}]$
 14. Projeter les matrices caractéristiques dans un nouveau sous espace: $Z_j = L^T \cdot Y_j$ pour $j = 1, \dots, M$
 15. Transformer les nouvelles matrices caractéristiques $Z_j (d_2 \times d_1)$ en vecteurs,
 16. Appliquer l'algorithme backpropagation détaillé dans le chapitre II
-

I.5 Approches locales bidimensionnelles : Yeux-ACP2D, Yeux-ADL2DoC et Yeux-ACPDL2D & Nez-ACP2D, Nez-ADL2DoC et Nez-ACPDL2D

I.5.1 Idée et méthodologie

Les trois méthodes globales précédentes (ACP2D, ADL2DoC et ACPDL2D) sont destinées à être appliquées sur le visage entier, par contre lorsqu'on dispose d'une base contenant des visages non occultés tandis qu'un ou plusieurs visages à identifier sont en partie cachés : les yeux peuvent être occultés par le port de lunettes, le nez par un cache-nez, la bouche par une barbe et/ou une moustache, ... (Figure IV.6) la reconnaissance basée sur ces visages sera sûrement erronée menant à des systèmes de reconnaissance de visages qui peuvent rejeter une personne légitime ou accepter un imposteur.



Figure IV.6. Occultations de certains éléments du visage : deux personnes de la base AR (1^{ère} ligne) et un extrait de la base ORL (2^{ème} ligne).

D'autres facteurs s'ajoutent au besoin d'identifier des personnes dont le visage est partiellement occulté à savoir:

- le système visuel humain, à des fins de reconnaissance de personnes, se base en premier lieu sur des informations globales (tout le visage) puis sur des informations locales (différents éléments du visage) pour affinement,
- le pouvoir discriminant des yeux [74] et
- la performance élevée des méthodes bidimensionnelles : l'ACP2D, l'ADL2DoC et l'ACPDL2D testées auparavant,

justifient notre idée qui consiste à utiliser, séparément, des éléments faciaux : les yeux, le nez et la bouche combinés avec les méthodes bidimensionnelles. Ainsi, nous corrigeons le problème des occultations partielles, des arrière plans, du style de coiffure et nous diminuons l'effet des variations de l'éclairage et la taille des données à traiter compte tenu que les régions contenant ces éléments sont de tailles réduites comparées à tout le visage.

Mais est-ce que la machine est capable de reconnaître une personne à partir de ses éléments faciaux pris séparément ? Et s'il est possible, quels sont les éléments contribuant à une bonne identification : les yeux, le nez ou la bouche ? Et est-ce que cette méthode de reconnaissance sera plus performante que les méthodes globales ?

I.5.2 Système proposé

Le principe des approches locales bidimensionnelles consiste à appliquer l'ACP2D, l'ADL2DoC et l'ACPDL2D (algorithme IV.1) sur deux régions locales du visage: une contenant les yeux (avec les sourcils) et l'autre le nez et effectuer ensuite la classification par un PMC d'où l'appellation « **Méthodes Locales Bidimensionnelles** ». Pour le nez, notons Nez-ACP2D, Nez-ADL2DoC et leur fusion Nez-ACPDL2D quant aux yeux, notons les approches : Yeux-ACP2D, Yeux-ADL2DoC et Yeux-ACPDL2D pour leur combinaison. Ainsi nous fusionnons deux grandes classes de méthodes de reconnaissance de visages; les approches globales avec celles basées sur les éléments faciaux.

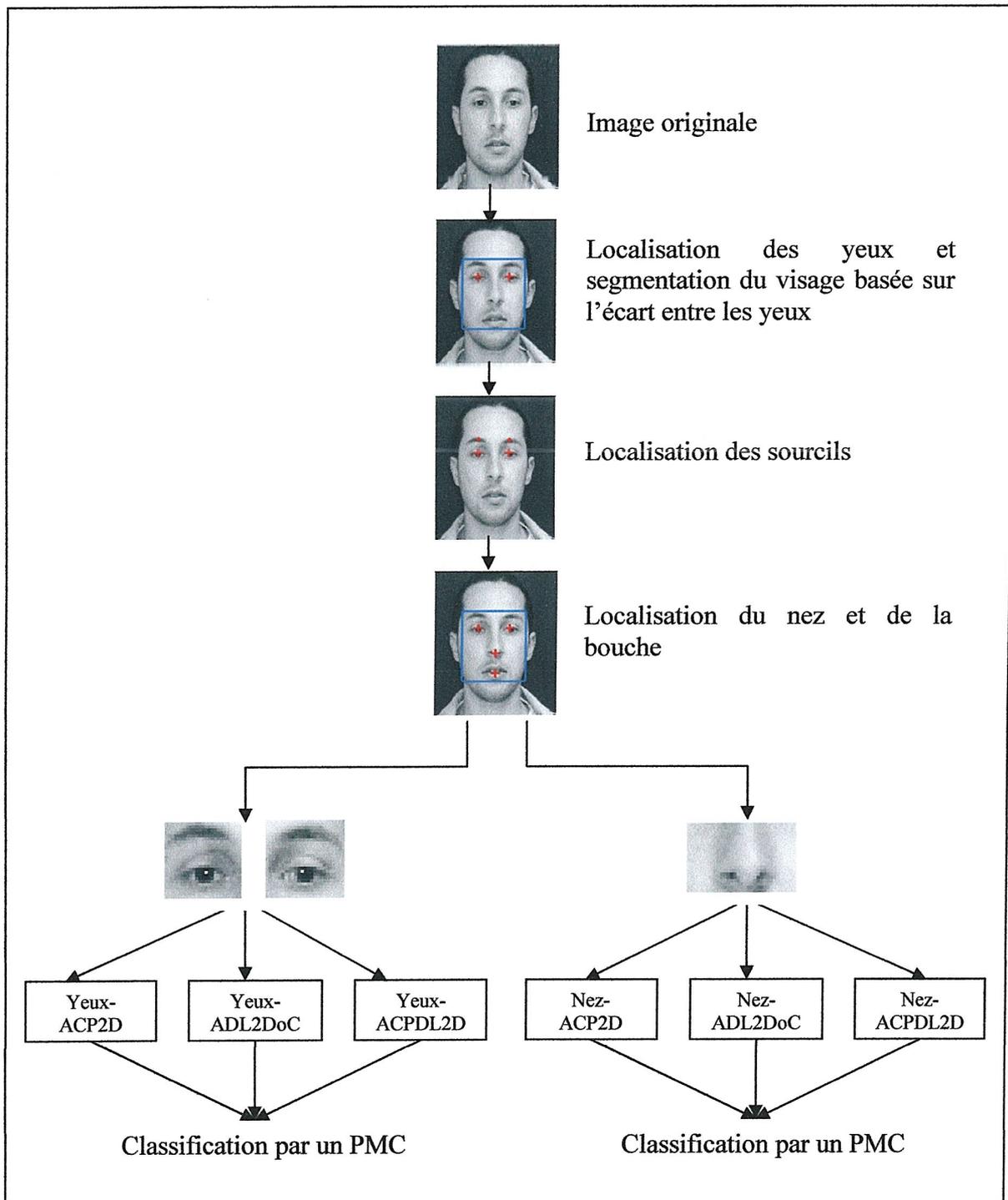


Figure IV.7. Principe des approches locales bidimensionnelles.

Ce système comporte différentes étapes ; elles seront détaillées dans les paragraphes suivants :

I.5.3 Localisation des éléments caractéristiques du visage: les yeux, les sourcils, le nez et la bouche

En raison de la prise de vues frontales et presque frontales des images de la séquence, et de la source d'éclairage, aussi frontale sur les iris, la méthode adoptée pour localiser les différents éléments du visage se base sur l'intensité de niveau de gris.

I.5.3.1 Localisation des iris

Pour localiser les iris et par conséquent les deux yeux [46], on effectue un balayage de toute l'image de haut en bas en vue de rechercher les maxima de la fonction d'intensité de niveau de gris ; les positions retournées sont celles des iris. L'application de cette méthode sur les images originales de la base B-LAIG, nous a permis de localiser précisément les iris des yeux comme le montre la figure IV.8.



Figure IV.8. Exemples de la localisation des sourcils et des yeux.

I.5.3.2 Localisation des sourcils

Pour les localiser, notre idée consiste à calculer la norme du gradient d'une image et rechercher, au dessus de chaque iris, la valeur maximale le long de la ligne verticale passant par l'iris localisé auparavant (figure IV.8).

I.5.3.3 Localisation du nez et de la bouche

Pour localiser le nez et la bouche, nous avons commencé par appliquer la méthodes de [46], cependant, son application à partir des images originales de la base B-LAIG, n'a pas donné de correctes localisations notamment pour les hommes moustachus et/ou barbus et les femmes portant un voile (les valeurs maximales peuvent coïncider avec une moustache, une barbe, un voile,...) comme l'illustre la figure IV.9. Commençons par rappeler le principe de la technique de [46] : pour pouvoir les localiser, on détermine la somme, en chaque point, de la norme du gradient spatio-temporel. L'image résultante vérifie:

$$I_{som} = \sum_{i=1}^{n_c} \sqrt{\left(\frac{\partial I_i(x,y,t)}{\partial x}\right)^2 + \left(\frac{\partial I_i(x,y,t)}{\partial y}\right)^2 + \mu \cdot \left(\frac{\partial I_i(x,y,t)}{\partial t}\right)^2} \quad (\text{IV.19})$$

Où $I(x,y,t)$ désigne l'intensité de niveau de gris, x et y les composantes spatiales, t la composante temporelle et μ un terme de normalisation dépendant de l'échantillonnage temporel de la séquence et du mouvement du visage. Ainsi, on prend en considération le fort gradient spatial de la région inférieure du nez et le gradient temporel du aux mouvements de la tête et des lèvres. Sur l'image I_{som} (équation IV.19), on recherche le long de la médiatrice du segment passant par les iris les deux régions ayant les plus fortes valeurs. Leurs positions correspondent à la partie inférieure du nez et au centre de la bouche. La figure IV.9 schématise des exemples de fausses localisations du nez et de la bouche à partir des images originales.

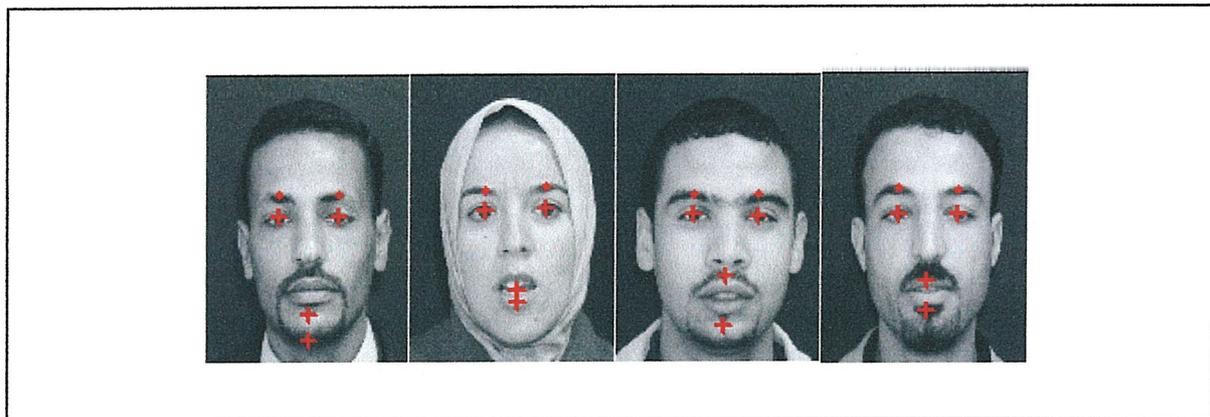


Figure IV.9. Exemples de fausses localisations du nez et de la bouche.

Pour améliorer la précision de la localisation, nous proposons une technique basée sur la partition et la géométrie du visage (figure IV.10). Pour cela, commençons par segmenter les visages en se basant sur les proportions anthropométriques entre les différents éléments faciaux. Connaissant l'écart entre les yeux, ces proportions, schématisée en figure IV.10(a), vérifient [94] :

$$H_{face} = 1.8 \cdot D_{eye} \tag{IV.20.a}$$

$$L_{face} = 2 \cdot W_{eye} + D_{eye} \tag{IV.20.b}$$

$$W_{eye} = 0.225 \cdot H_{face} \tag{IV.20.c}$$

$$H_{eye} = \frac{1}{5} H_{face} \tag{IV.20.d}$$

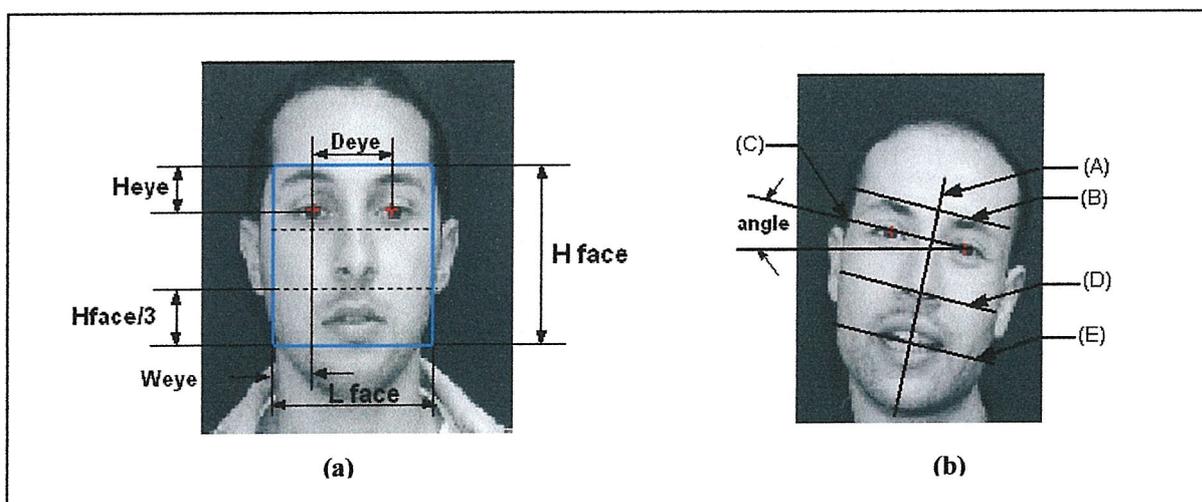


Figure IV.10. Partitionnement (a) et géométrie (b) du visage.

Pratiquement, nous avons utilisé, au lieu de la valeur 1.8 de l'équation (IV.20.a) la valeur «2.1»

Une fois ces proportions déterminées, le visage peut être partitionné en 3 trois parties égales, la première contient les sourcils et les yeux, la seconde contient le nez et la troisième la bouche. Pour améliorer la précision, nous limitons la recherche par la fin de la 2^{ème} partition pour le nez, et la fin de la 3^{ème} partie pour la bouche. Pour les visages orientés, la précision est améliorée grâce à la géométrie du visage (voir figure IV.10(b)): la ligne (C) passant par les iris est toujours perpendiculaire à la ligne (A) passant par le nez et la bouche. Les lignes (B), (D) et (E) qui passent respectivement par les sourcils, le nez et la bouche sont parallèles à la ligne (C). Le nez est donc le point d'intersection de la ligne (A) et (D), la bouche est localisée par l'intersection des lignes (A) et (E).

La segmentation combinée avec la géométrie du visage nous a permis d'améliorer la précision de la localisation. A cause de l'activité labiale, la bouche est localisée par un point situé soit au centre soit sur l'une des lèvres. La figure IV.11 schématise des exemples de localisation des caractéristiques faciales.

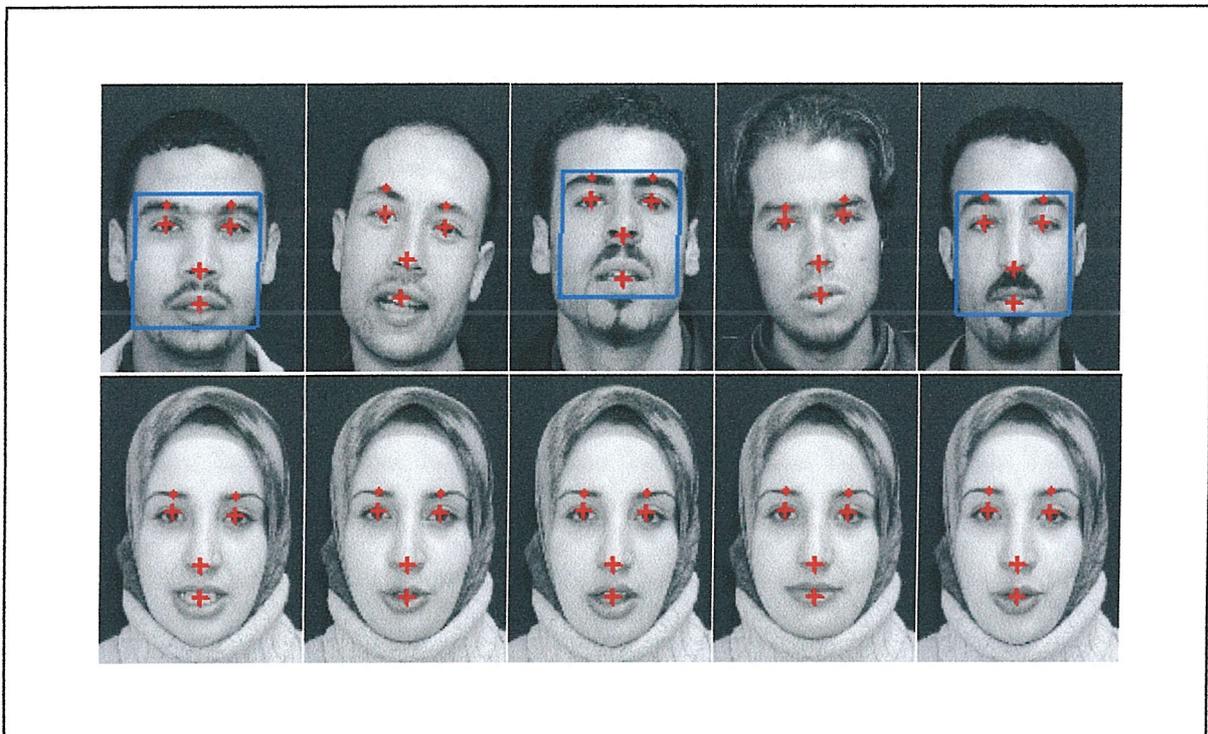


Figure IV.11. Exemples de correctes localisations des éléments du visage : différentes personnes (1^{ère} ligne) et 5 images d'une séquence (2^{ème} ligne).

I.5.4 Extraction des régions contenant les éléments du visage

Une fois, les sourcils et les iris localisés, l'extraction de régions qui contiennent ces éléments, revient à extraire pour chaque visage de la séquence deux rectangles de tailles suffisantes : quelques pixels au dessus des sourcils et quelques pixels au dessous des iris tout en veillant à ne pas prendre les foulards des femmes. Ainsi, ces rectangles ne contiennent que le sourcil et l'œil gauche et le sourcil et l'œil de droite. Un extrait de ces régions juxtaposées est donné en figure IV.12. Il est clair qu'il y a des variations inter-classes importantes dues à la fois à la forme des sourcils, à la forme des yeux et à leurs mouvements.

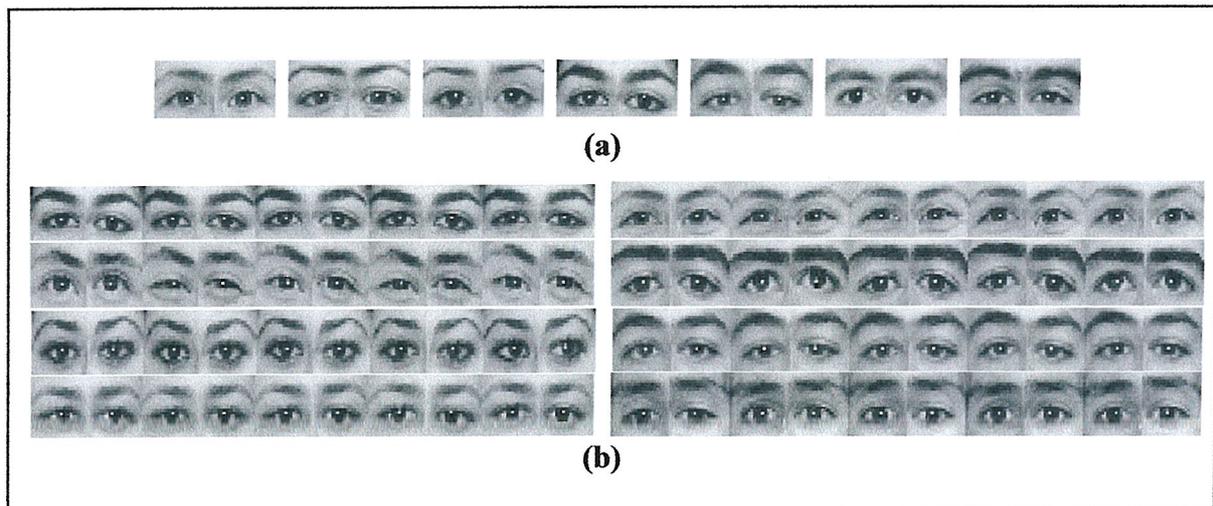


Figure IV.12. Exemples de régions contenant les yeux et les sourcils :
(a) une seule image et (b) concaténation de 5 images.

Pour les nez, nous avons veillé à ce que les régions ne contiennent que les nez sans les moustaches, pour assurer une correcte reconnaissance de personnes basée uniquement sur la forme des nez compte tenu que les hommes peuvent laisser ou enlever leurs moustaches et barbes. Les régions extraites sont de tailles (21×33) .

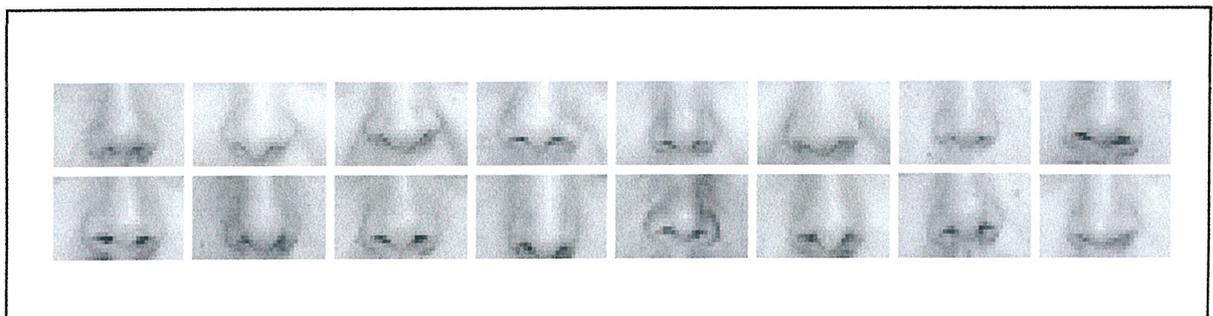


Figure IV.13. Exemples de régions contenant les nez.

Nous avons remarqué que la région de la bouche présente d'importantes variations intra-classes dues aux mouvements des lèvres lors de la prononciation d'une phrase lors de l'acquisition des images de visages (figure IV.14).

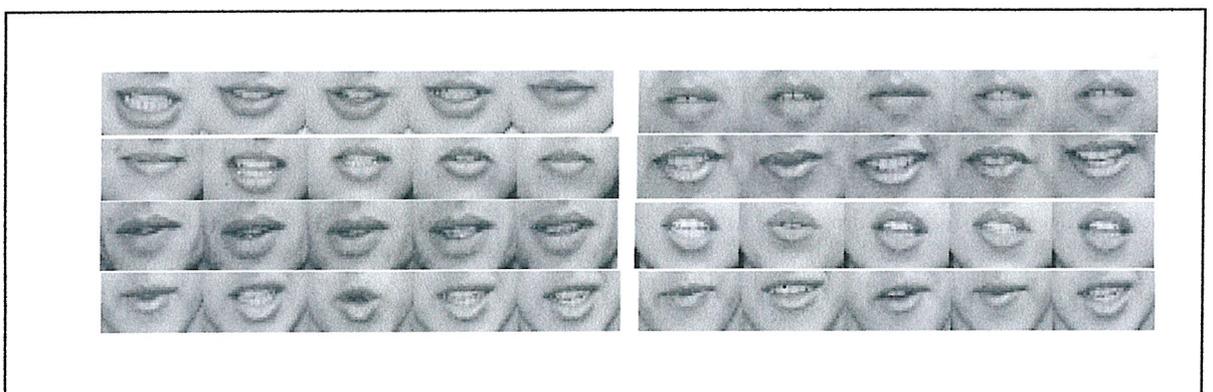
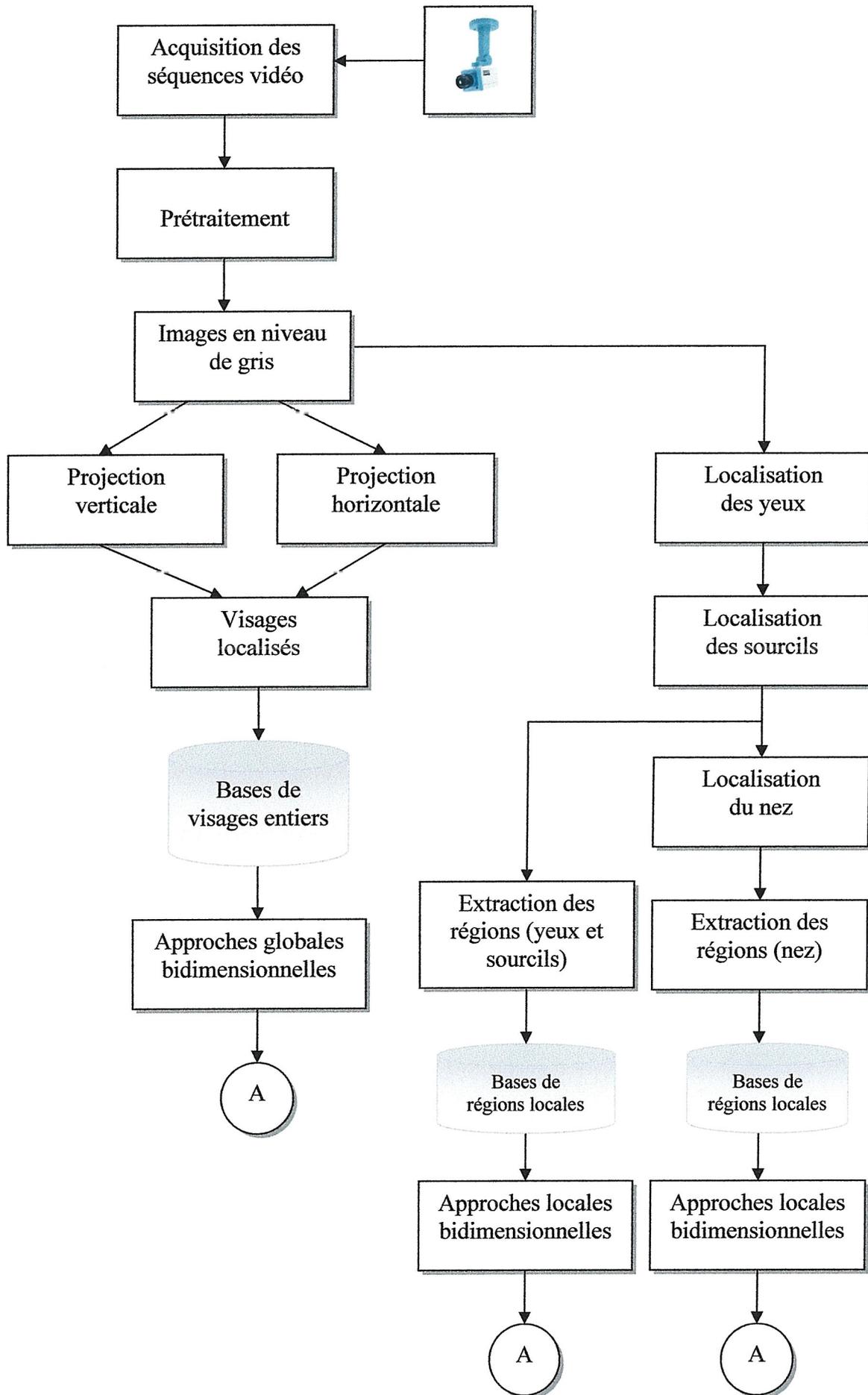
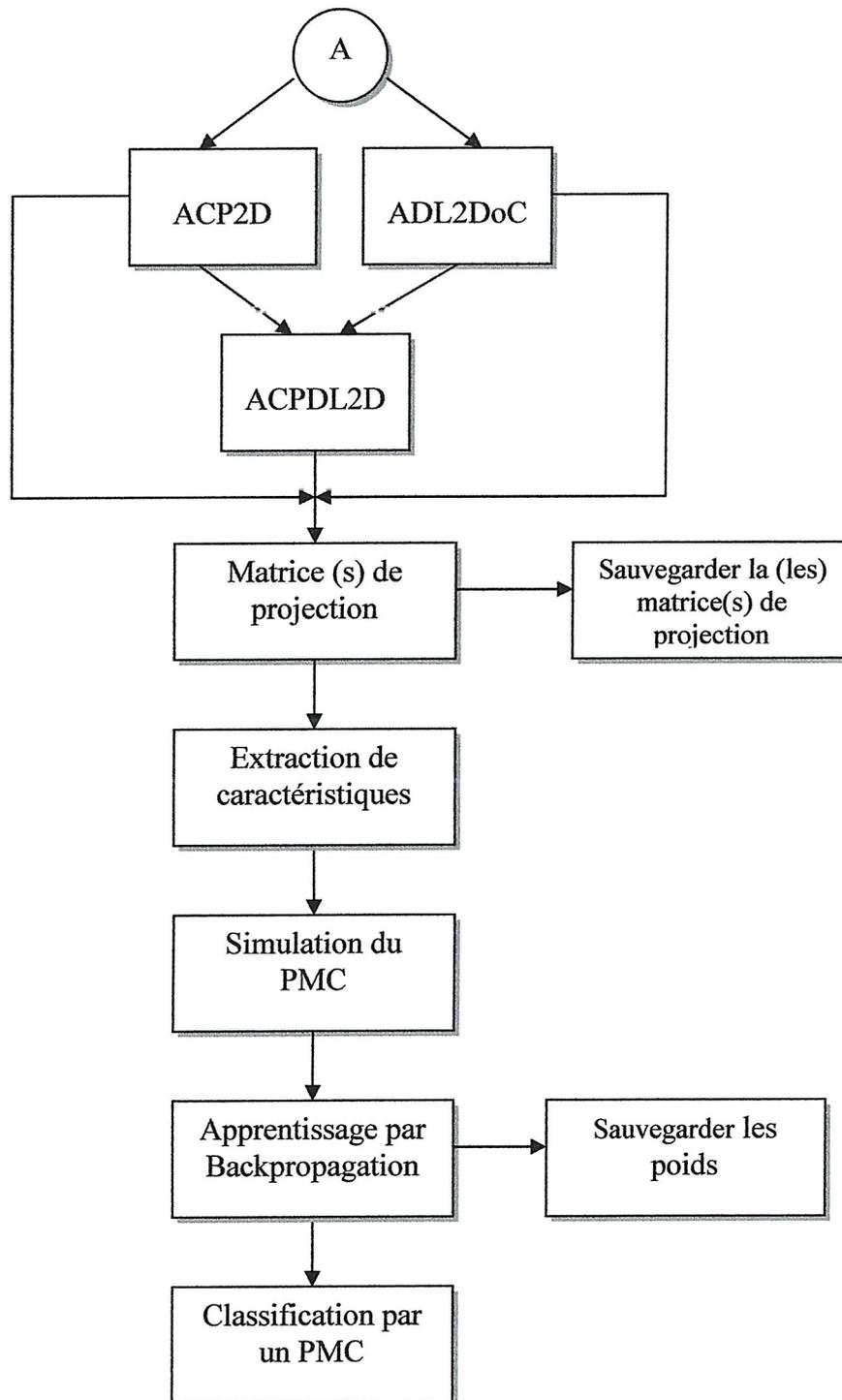


Figure IV.14. Exemples des mouvements des lèvres de personnes prononçant la phrase «reconnaissance automatique de visages». (4^{ème} ligne: la bouche d'une personne).

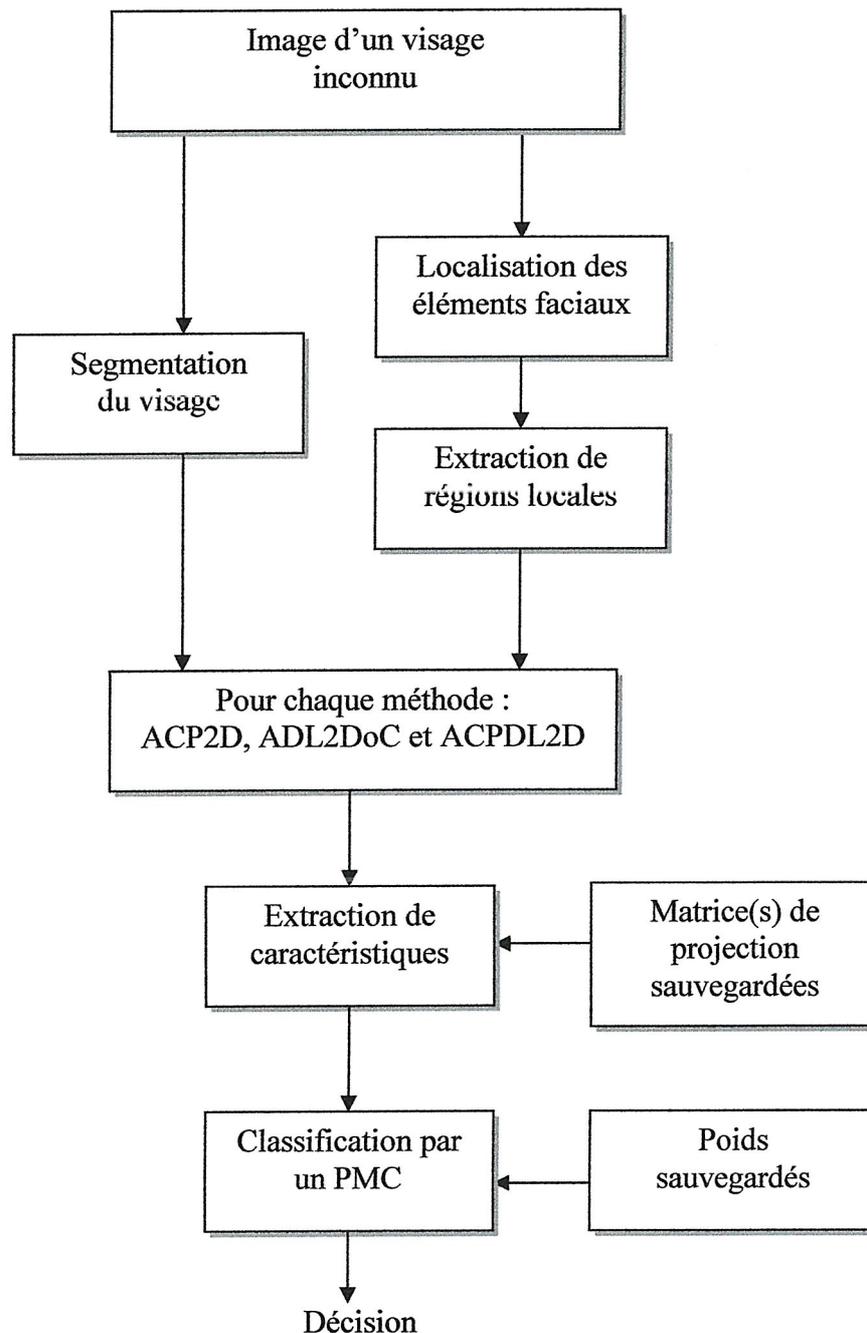
II. Résultats Expérimentaux et Interprétations

L'organigramme suivant récapitule le travail présenté dans le cadre de cette recherche:





Organigramme IV.2. *Etapes de la phase d'apprentissage du système de reconnaissance faciale.*



Organigramme IV.3. *Etapes de la phase de test du système de reconnaissance faciale.*

II.1 Résultats expérimentaux des méthodes globales

Pour mettre en valeur les performances du système proposé, nous avons mené des tests sur la base de visages du LAIG « B-LAIG » en appliquant l'ACP2D, l'ADL2DoC et l'approche hybride proposée ACPDL2D avec le classificateur neuronal.

Les résultats obtenus par les trois méthodes, sont évalués en utilisant le Hold Out qui consiste à partitionner aléatoirement la base de visages en deux sous bases de 200 images

chacune, l'une pour l'apprentissage et l'autre pour le test. Cette opération est répétée six fois avant de moyenner les résultats.

II.1.1 Présentation de la base de visages utilisée

La base de visages utilisée a été acquise au LAIG sous des conditions d'éclairage contrôlées et avec des prises de vues frontales et presque frontales mais avec des variations dans l'échelle. Cette base contient 400 images de 40 personnes (20 hommes et 20 femmes) ayant chacun une séquence de 10 images. Lors de l'acquisition des images, chaque personne est invitée à prononcer la phrase « reconnaissance automatique de visages ».



Figure IV.15. Extrait de la base B-LAIG.

II.1.2 Résultats de la localisation automatique de visages

En appliquant la méthode de la localisation automatique expliquée dans le paragraphe I.4.2 de ce chapitre, un extrait de visages localisés est donné dans la figure ci-dessous :

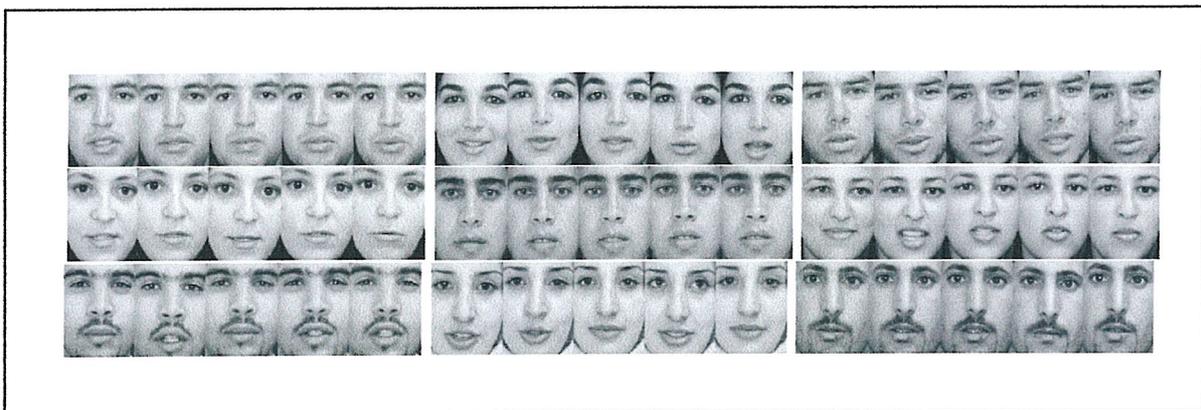


Figure IV.16. Extrait d'images de visages après leur localisation.

II.1.3 Effet du choix de la topologie du PMC

Les performances du PMC dépendent de l'initialisation des poids, du taux (pas) d'apprentissage et du nombre de neurones de la couche cachée, c'est pourquoi nous avons réalisé plusieurs tests pour chaque méthode avant de préserver une topologie donnant la meilleure performance (l'apprentissage est arrêté, pour tous nos tests, lorsque l'erreur quadratique moyenne est ≤ 0.01). Les taux de reconnaissance reportés sont la moyenne des

taux obtenus sur les six partitions de test pour chaque méthode et pour chaque nombre de vecteurs propres retenus. L'architecture du PMC est N-H-J avec :

- N : nombre de neurones de la couche d'entrée. Ce dernier diffère selon l'approche appliquée; $(n \times d_1)$ neurones pour l'ACP2D, $(d_2 \times m)$ pour l'ADL2DoC et $(d_2 \times d_1)$ neurones pour l'ACPDL2D,
- H : nombre de neurones de la couche cachée. Il est déterminé empiriquement,
- J : nombre de neurones de la couche de sortie correspondant au nombre de classes de la base de visages ($J = 40$).

II.1.4 Reconnaissance de visages par les méthodes globales

Les méthodes bidimensionnelles n'échappent pas au problème du choix du nombre de valeurs propres à préserver. Ce choix dépend de la méthode appliquée et de la base de visages utilisée comme le montre les expériences suivantes menées en fonction de différents nombres de vecteurs propres.

II.1.4.1 Résultats de Neuro-ACP2D

Pour tester l'effet du nombre de valeurs propres sur l'ACP2D, nous avons réalisé des expériences avec des valeurs propres (tableau IV.1) correspondantes respectivement à 53.72%, 63.03%, 70.87%, 85.49%, 88.69% et 90.89% de la variance totale (obtenues pour la 1^{ère} partition de la base B-LAIG) et un pas d'apprentissage $\eta = 0.1$. Le meilleur résultat est obtenu pour 8 valeurs propres et donc une taille de l'image caractéristique (73×8).

Neuro-ACP2D						
H	25	25	30	35	37	38
$n \times d_1$	73×2	73×4	73×6	73×8	73×10	73×12
Taux (%)	95.50	96.16	96.00	96.50	96.33	96.16

Tableau IV.3. Taux de reconnaissance obtenus par l'ACP2D en fonction de différents nombres de vecteurs propres.

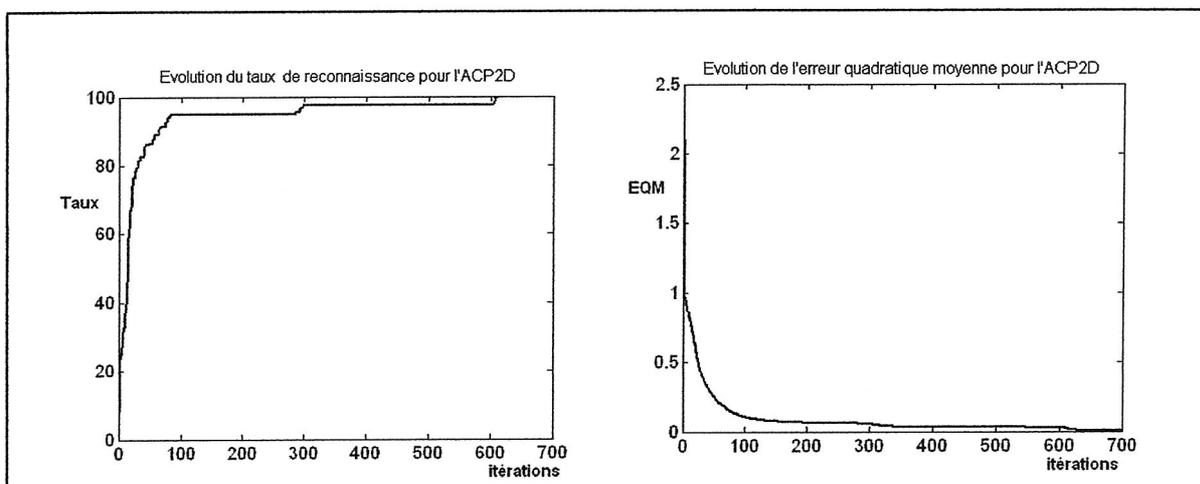


Figure IV.17. Evolution du taux de reconnaissance (à gauche) et de l'erreur quadratique moyenne (à droite) obtenus par l'ACP2D ($d_1=4$) sur l'ensemble d'apprentissage.

II.1.4.2 Résultats de Neuro-ADL2DoC

Pour cette méthode, le meilleur résultat est obtenu pour 6 vecteurs propres, un pas d'apprentissage $\eta = 0.1$ et donc une image caractéristique de taille (6×56) . On constate aussi que l'ADL2DoC est plus performante que l'ACP2D comme le montre le tableau IV.4.

Neuro-ADL2DoC						
H	27	29	29	31	30	34
$d_2 \times m$	2×56	4×56	6×56	8×56	10×56	12×56
Taux(%)	95.66	96.75	97.25	96.91	97.00	96.66

Tableau IV.4. Taux de reconnaissance obtenus par l'ADL2DoC en fonction de différents nombres de vecteurs propres.

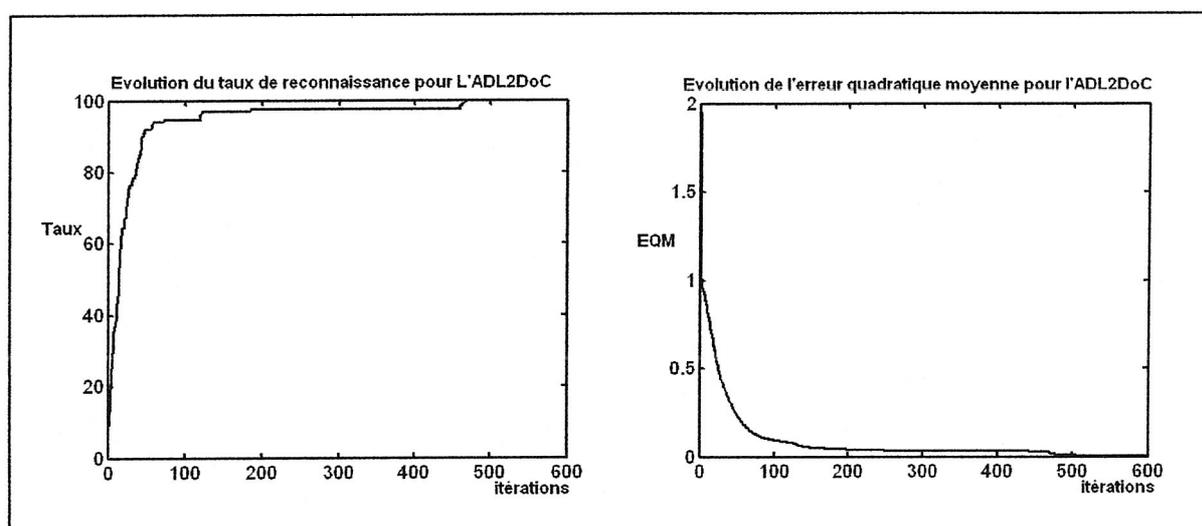


Figure IV.18. Evolution du taux de reconnaissance (à gauche) et de l'erreur quadratique moyenne (à droite) obtenus par l'ADL2DoC ($d_2=10$) sur l'ensemble d'apprentissage.

II.1.4.3 Résultats de Neuro-ACPDL2D

Les tests effectués en appliquant l'ACP2D neuronale montrent qu'elle est optimale pour 8 vecteurs propres, c'est pourquoi nous avons préservé ce nombre pour notre approche proposée Neuro-ACPDL2D tout en changeant le nombre de vecteurs propres pour l'ADL2DoC. Le tableau IV.5 illustre les résultats obtenus qui dépassent ceux de l'ACP2D et l'ADL2DoC en termes d'espace mémoire et sont supérieurs à ceux obtenus avec l'ACP2D (sauf pour le cas de 2 valeurs propres) et en concurrence avec ceux obtenus avec l'ADL2DoC.

Neuro-ACPDL2D						
H	27	24	25	34	36	37
$d_2 \times d_1$	2×8	4×8	6×8	8×8	10×8	12×8
Taux(%)	95.33	96.50	96.91	97.00	97.00	96.75

Tableau IV.5. Taux de reconnaissance obtenus par Neuro-ACPDL2D en fonction de différents nombres de vecteurs propres.

I.1.5 Comparaison des résultats

L'approche Neuro-ACPDL2D permet de reconnaître **97%** de la base de visages B-LAIG avec seulement **64 coefficients** au lieu de $73 \times 56 = 4088$ pixels en utilisant tout le visage, 584 coefficients avec l'ACP2D et 336 coefficients avec l'ADL2DoC. La compression de données originales est donc très importante :

Méthodes	ACPDL2D	ADL2DoC	ACP2D
Taux de compression (%)	98.43	91.78	85.71

Tableau IV.6. Comparaison des taux de compression de données.

Il est important de noter que pour la même base de visages et avec le même type de classificateur (PMC), les trois méthodes globales sont meilleures que l'approche géométrique dite Neuro-ACP proposée dans [69] comme le présente le tableau IV.7.

Méthodes	Taille	Taux (%)
Neuro-ACP [69]	-----	94.50
Neuro-ACP2D	$73 \times 8 = 584$	96.50
Neuro-ADL2DoC	$6 \times 56 = 336$	97.25
Neuro-ACPDL2D	$8 \times 8 = 64$	97.00

Tableau IV.7. Comparaison entre les méthodes globales et la méthode géométrique.

II.2 Reconnaissance de visages par les méthodes locales bidimensionnelles

Pour toutes les expériences menées, la base d'apprentissage est partitionnée aléatoirement en deux bases, l'une servant pour l'apprentissage et l'autre pour le test. Nous avons répété cette opération six fois avant de moyenniser les taux de reconnaissance obtenus sur les bases de test.

II.2.1 Applications sur les régions contenant les sourcils et les yeux

II.2.1.1 Résultats de Yeux-ACP2D

La reconstruction des régions des yeux en appliquant Yeux-ACP2D est illustrée en figure IV.19 :

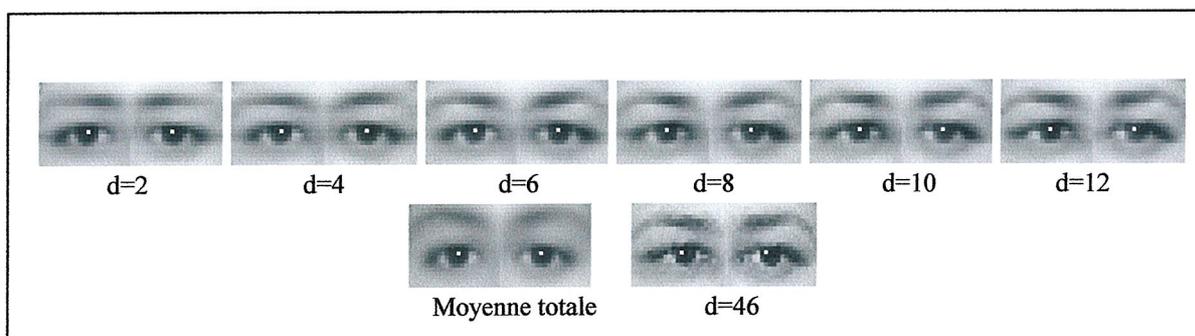


Figure IV.19. Exemples de reconstruction d'une région contenant des yeux en fonction des vecteurs propres retenus.

et les résultats des tests sont résumés dans le tableau suivant :

Yeux-ACP2D						
H	15	16	18	19	19	18
$n \times d_1$	21×2	21×4	21×6	21×8	21×10	21×12
Taux(%)	96.58	97.83	98.41	99.25	99.16	99.16

Tableau IV.8. Taux de reconnaissance obtenus par Yeux-ACP2D en fonction de différents nombres de vecteurs propres retenus.

II.2.1.2 Résultats de Yeux-ADL2DoC

Les résultats de l'application de la méthode Yeux-ADL2DoC sont donnés dans le tableau IV.9. Le meilleur résultat est obtenu pour 8 vecteurs propres et un pas d'apprentissage $\eta = 0.1$

Yeux-ADL2DoC						
H	12	11	14	18	20	21
$d_2 \times m$	2×46	4×46	6×46	8×46	10×46	12×46
Taux(%)	97.66	98.83	98.75	99.25	99.16	99.25

Tableau IV.9. Taux de reconnaissance obtenus par Yeux-ADL2DoC en fonction de différents nombres de vecteurs propres retenus.

II.2.1.3 Résultats de Yeux-ACPDL2D

La fusion de Yeux-ACPL2D et de Yeux-ADL2DoC donne Yeux-ACPDL2D qui est, comme le montre le tableau IV.10, optimale pour 8 vecteurs propres. Sa supériorité est claire notamment pour les cas de 2, 4 et 6 vecteurs propres.

Yeux-ACPDL2D						
H	13	16	17	21	23	24
$d_1 \times d_2$	2×8	4×8	6×8	8×8	10×8	12×8
Taux(%)	98.41	99.08	99.25	99.41	99.41	99.33

Tableau IV.10. Taux de reconnaissance obtenus par Yeux-ACPDL2D en fonction de différents nombres de vecteurs propres retenus.

II.2.1.4 Comparaison des résultats

Le tableau ci-dessous récapitule les résultats des trois méthodes précédentes dont l'approche optimale est **Yeux-ACPDL2D** suivie de Yeux-ADL2DoC et de Yeux-ACP2D.

Méthodes	Taille	Taux (%)
Yeux-ACPDL2D	8×8	99.41
Yeux-ADL2DoC	8×46	99.25
Yeux-ACP2D	21×8	99.25

Tableau IV.11. Comparaison entre les méthodes locales bidimensionnelles appliquées sur les régions contenant les sourcils et les yeux.

II.2.2 Applications sur les régions contenant les nez

II.2.2.1 Résultats de Nez-ACP2D

La reconstruction des régions des nez est illustrée par l'exemple de la figure ci-dessous. Nous avons effectué plusieurs tests pour lesquels 12 valeurs propres correspondent à 95.10% de l'information totale calculée sur la première partition de la base.

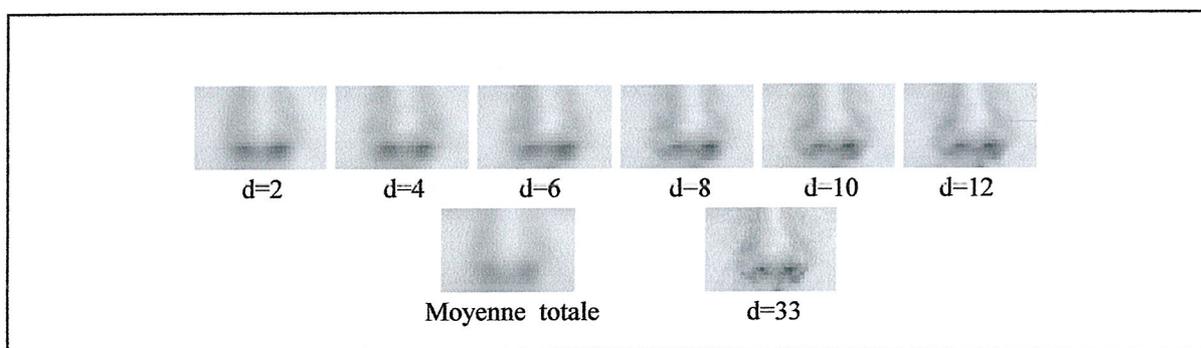


Figure IV.20. Exemples de reconstruction d'une région contenant un nez en fonction des vecteurs propres retenus.

Le tableau IV.12 résume les résultats montrant que l'approche Nez-ACP2D est optimale pour 10 valeurs propres arrivant à un taux de reconnaissance de 97.08%. ($\eta = 0.3$).

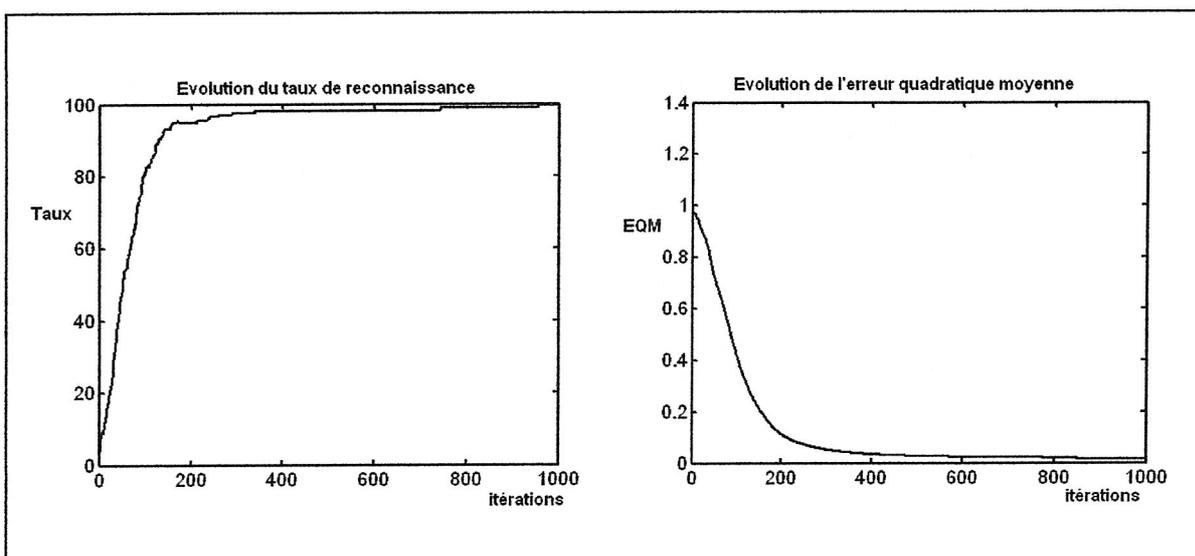


Figure IV.21. Evolution du taux de reconnaissance (à gauche) et de l'erreur quadratique moyenne (à droite) obtenus sur la base d'apprentissage par Nez-ACP2D.

Nez-ACP2D						
H	19	22	22	24	25	27
n × d₁	2 × 21	21 × 4	21 × 6	21 × 8	21 × 10	21 × 12
Taux (%)	92.33	95.91	96.66	96.83	97.08	97.08

Tableau IV.12. Taux de reconnaissance obtenus par Nez-ACPDL2D en fonction du nombre de vecteurs propres retenus.

II.2.2.2 Résultats de Nez-ADL2DoC

D'après les résultats obtenus, l'approche Nez-ADL2DoC donne le meilleur taux de reconnaissance pour 4 vecteurs propres seulement ($\eta = 0.3$).

Nez-ADL2DoC						
H	15	18	23	22	26	28
d₂ × m	2 × 33	4 × 33	6 × 33	8 × 33	10 × 33	12 × 33
Taux (%)	97.16	97.75	97.25	97.08	97.00	97.16

Tableau IV.13. Taux de reconnaissance obtenus par Nez-ADL2DoC en fonction du nombre de vecteurs propres retenus.

II.2.2.3 Résultats de l'approche hybride Nez-ACPDL2D

L'approche hybride Nez-ACPDL2D est optimale pour seulement 40 coefficients, arrivant à reconnaître **98.16%** de la base de visages. Pour tous les cas de vecteurs propres retenus, l'approche Nez-ACPDL2D dépasse les autres résultats notamment ceux de l'approche Nez-ACP2D comme l'illustre le tableau suivant :

Nez-ACPDL2D						
H	19	19	20	23	25	26
d₁ × d₂	2 × 10	4 × 10	6 × 10	8 × 10	10 × 10	12 × 10
Taux (%)	97.25	98.16	97.66	97.75	97.41	97.41

Tableau IV.14. Taux de reconnaissance obtenus par Nez-ACPDL2D en fonction du nombre de vecteurs propres retenus.

II.2.2.4 Comparaison des résultats

Le tableau IV.15 montre la supériorité de l'approche hybride Nez-ACPDL2D sur les autres méthodes : Nez-ACP2D et Nez-ADL2DoC en termes d'espace mémoire et de taux de reconnaissance.

Méthodes	Taille	Taux (%)
Nez-ACPD2D	4×10	98.16
Nez-ADL2DoC	4×33	97.75
Nez-ACP2D	21×10	97.08

Tableau IV.15. Comparaison entre les méthodes locales bidimensionnelles appliquées sur les régions contenant les nez.

II.2.3 Comparaison entre les résultats obtenus par les méthodes bidimensionnelles globales et locales

Le tableau ci-dessous récapitule tous les résultats obtenus en appliquant les méthodes bidimensionnelles globales et locales avec le PMC.

Méthodes	Méthodes locales bidimensionnelles				Méthodes Globales	
	Sourcils et yeux		Nez		Tout le visage	
	taille	Taux (%)	taille	Taux (%)	taille	Taux (%)
ACP2D	21×8	99.25	21×10	97.08	73×8	96.50
ADL2DoC	8×46	99.25	4×33	97.75	6×56	97.25
ACPD2D	10×8	99.41	4×10	98.16	8×8	97.00

Tableau IV.16. Comparaison des résultats.

Conclusion

Le tableau ci-dessous récapitule les avantages de l'ACP2D et de l'ADL2D ainsi que celles de l'approche hybride Neuro-ACPD2D qui bénéficie de leur fusion avec le PMC.

ACP2D	ADL2DoC	Neuro-ACPD2D
<ul style="list-style-type: none"> - Elimination des redondances au niveau des lignes, - Projection unilatérale, - Réduction du temps de calcul de la matrice de covariance, - Pouvoir représentatif, - Plus proche voisin. 	<ul style="list-style-type: none"> - Elimination des redondances au niveau des colonnes, - Projection unilatérale, - Correction du problème de la singularité, - Pouvoir discriminant, - Plus proche voisin. 	<ul style="list-style-type: none"> - Elimination des redondances au niveau des lignes et des colonnes, - Projection bilatérale, - Applicable pour les problèmes de petites tailles, - Pouvoir représentatif et discriminant, - Pouvoir de généralisation du PMC.

Tableau IV.17. Avantages des méthodes bidimensionnelles.

Les résultats de simulations nous permettent de conclure que :

- les méthodes locales bidimensionnelles sont plus performantes que les trois méthodes globales. Les yeux sont les plus discriminants suivis du nez et du visage entier,
- l'approche hybride Neuro-ACPD2D globale dépasse les deux approches : Neuro-ACP2D et Neuro-ADL2DoC en termes d'espace mémoire. Son taux de reconnaissance dépasse celui de Neuro-ACP2D globale,
- l'approche hybride Neuro-ACPD2D locale dépasse, en termes d'espace mémoire et de taux de reconnaissance, Neuro-ACP2D et Neuro-ADL2DoC locales. Cette supériorité est claire notamment pour les cas de 2,4 et 6 vecteurs propres,
- le nez permet tout seul, par Neuro-ACPD2D, de reconnaître 98.16% de la base de visages B-LAIG.

En conclusion, l'approche Neuro-ACPD2D avec ses deux versions globale et locale, comparée à Neuro-ACP2D et Neuro-ADL2DoC est très performante. La supériorité des approches locales bidimensionnelles peut être justifiée par les performances des approches bidimensionnelles et le pouvoir discriminant des éléments faciaux.

Conclusion Générale

Le travail présenté dans ce mémoire s'inscrit dans le contexte de l'identification automatique de visages à partir de séquences d'images vidéo. Pour cela, la base de visages utilisée contient 400 images de 40 personnes de sexes différents. Elle a été acquise au LAIG en 2003/2004 sous des conditions d'illumination contrôlées, avec des vues frontales et presque frontales et des expressions faciales dues à la prononciation de la phrase « reconnaissance automatique de visages ».

Le visage est une caractéristique biométrique non intrusive et très acceptable par les utilisateurs. Ce dernier est le cœur des systèmes de reconnaissance faciale qui sont des systèmes typiques de reconnaissance de formes. Ils comportent donc deux étapes : l'une pour l'extraction de caractéristiques (signatures) et l'autre pour la classification de celles-ci.

A cet objectif, la 1^{ère} partie de notre contribution a été dévolue à la proposition d'une approche hybride bidimensionnelle Neuro-ACPD2D. Le principe de cette méthode globale, consiste, après la localisation automatique de visages, à effectuer l'étape d'extraction de caractéristiques par la fusion de deux méthodes récemment proposées: l'ACP2D et l'ADL2DoC et ensuite la classification des caractéristiques obtenues par un PMC. Ainsi via la fusion proposée, cette approche contourne les problèmes du temps de calcul et de la singularité de la matrice de covariance, et offre un codage et une compression de données très efficaces avec une identification satisfaisante. Cela peut être justifié par les pouvoirs représentatif de l'ACP2D, discriminant de l'ADL2D et de généralisation du PMC.

Nous proposons dans la 2^{ème} partie, des approches locales bidimensionnelles au sens où elles s'appliquent sur des régions faciales locales: les yeux (avec les sourcils) et le nez. Ainsi l'identification est basée uniquement sur la forme de ces éléments qui combinée avec les approches globales bidimensionnelles bénéficie des performances de ces dernières et de la discrimination des éléments du visage. Comparées aux approches globales et géométrique, les approches locales bidimensionnelles sont dédiées à l'identification de personnes dont le visage est en partie cachée et sont privilégiées, car elles traitent des régions de tailles réduites, invariantes aux arrières plans, au port de lunettes, de cache-nez, aux moustaches et moins sensibles aux variations d'illumination et de la couleur de peau. Les résultats montrent, non seulement, que les yeux sont les plus discriminants suivis du nez et du visage entier mais que le nez, qui n'a pas suscité l'intérêt des chercheurs permet, à partir des vues de face, d'identifier efficacement une personne ce qui ouvre une voie de recherche basée sur cette caractéristique faciale relativement invariante aux expressions faciales et à l'âge.

Les approches proposées peuvent être appliquées pour l'identification de suspects à partir de leurs photos d'identité et de passagers dans les aéroports. En perspective, nous suggérons d'étudier, le pouvoir discriminant des yeux et du nez à partir des vues de profil en appliquant l'approche ACPDL2D locale sur des bases de visages internationales et avec d'autres types de classificateurs.

Annexe A

I.1 Présentation de la base de données IRIS de Fisher

Parmi les bases de données utilisées pour l'évaluation des classificateurs, la base de données IRIS de FISHER [2]. Elle est constituée de trois classes: Setosa, Versicolor et Virginica, de 50 échantillons de fleurs chacune, ainsi dans sa totalité, elle en contient 150. Le vecteur caractéristique de chaque échantillon contient 4 attributs: longueur et largeur de la sépale ainsi que longueur et largeur de la pétale.



Setosa



Versicolor



Virginica

I.2 La base 10-5-10 (The 10-5-10 encoder problem)

Cette base de données est très utilisée pour tester les algorithmes et les techniques proposés en classification. Elle a été utilisée par Fahlman.88 [13] pour étudier l'effet du choix des paramètres du réseau entraîné par l'algorithme Backpropagation avec moment sur la convergence ainsi que pour tester son algorithme Quickprop. Riedmiller.93 [21] l'a également utilisée pour tester son algorithme RPROP.

Le problème d'encodeur 10-5-10 est formé de 10 exemples et 10 sorties binaires. Chaque exemple est formé de 10 composantes : une vaut 1 et les autres 0. Le réseau de neurones correspondant comporte ainsi 10 neurones dans la couche d'entrée, 10 autres dans la couche de sortie et 5 neurones dans la couche cachée.

Exemples d'entrée	Sorties désirées
1 0 0 0 0 0 0 0 0 0	1 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0 0 0	0 1 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0	0 0 1 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0	0 0 0 1 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0	0 0 0 0 1 0 0 0 0 0
0 0 0 0 0 1 0 0 0 0	0 0 0 0 0 1 0 0 0 0
0 0 0 0 0 0 1 0 0 0	0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 0 0 1 0 0	0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 0 1 0	0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 0 1	0 0 0 0 0 0 0 0 0 1

Tableau A.1 Base de données 10-5-10.

Annexe B

I. Détermination des dérivées des sigmoïdes

a) Pour la couche de sortie

$$z_j = g(s_j) = \frac{1}{1 + e^{-\alpha s_j}}$$

$$g'(s_j) = \alpha e^{-\alpha s_j} \cdot \frac{1}{(1 + e^{-\alpha s_j})^2} = \alpha (1 + e^{-\alpha s_j} - 1) z_j^2$$

$$g'(s_j) = \alpha (z_j^{-1} - 1) z_j^2 = \alpha z_j (1 - z_j)$$

b) Pour la couche cachée

$$y_m = f(r_m) = \frac{1}{1 + e^{-\alpha r_m}} \quad \text{et} \quad f'(r_m) = \alpha y_m (1 - y_m)$$

II. Détermination des dérivées partielles de la fonction erreur

a) Pour les poids et les biais de la couche de sortie

$$\begin{aligned} \frac{\partial E}{\partial u_{mj}} &= \left(\frac{\partial E}{\partial z_j} \right) \left(\frac{\partial z_j}{\partial s_j} \right) \left(\frac{\partial s_j}{\partial u_{mj}} \right) = \left[\frac{\partial}{\partial z_j} \left(\sum_{j=1}^J (d_j - z_j)^2 \right) \right] \left[\frac{\partial}{\partial s_j} g(s_j) \right] \left[\frac{\partial}{\partial u_{mj}} \left(\sum_{m=1}^M y_m u_{mj} + b_j^o \right) \right] \\ &= -2 (t_j - z_j) \alpha z_j (1 - z_j) y_m \end{aligned}$$

$$\frac{\partial E}{\partial b_j^o} = -2 (t_j - z_j) \alpha z_j (1 - z_j)$$

b) Pour les poids et les biais de la couche cachée

$$\begin{aligned} \frac{\partial E}{\partial w_{nm}} &= \left(\frac{\partial E}{\partial y_m} \right) \left(\frac{\partial y_m}{\partial r_m} \right) \left(\frac{\partial r_m}{\partial w_{nm}} \right) \\ &= \left[\frac{\partial}{\partial y_m} \left(\sum_{j=1}^J (d_j - z_j)^2 \right) \right] \left[\frac{\partial}{\partial r_m} f(r_m) \right] \left[\frac{\partial}{\partial w_{nm}} \left(\sum_{n=1}^N w_{nm} x_n + b_m^H \right) \right] \\ &= \left[\frac{\partial}{\partial s_j} \left(\sum_{j=1}^J (d_j - z_j)^2 \right) \cdot \frac{\partial s_j}{\partial y_m} \right] f'(r_m) x_n \\ &= \left[-2 \sum_{j=1}^J (d_j - z_j) g'(s_j) \cdot \frac{\partial}{\partial y_m} \left(\sum_{m=1}^M u_{mj} y_m + b_j^o \right) \right] f'(r_m) x_n \\ &= -2 \left(\sum_{j=1}^J (d_j - z_j)^2 \alpha z_j (1 - z_j) u_{mj} \right) \alpha y_m (1 - y_m) x_n \\ \frac{\partial E}{\partial b_m^H} &= -2 \left(\sum_{j=1}^J (d_j - z_j)^2 \alpha z_j (1 - z_j) u_{mj} \right) \alpha y_m (1 - y_m) \end{aligned}$$

Annexe C

$X(n \times m)$: matrice de visage.

\bar{X} : matrice moyenne de toutes les images de visages de la base d'apprentissage.

M : nombre d'images.

C : nombre de classes.

n_c : nombre d'images de la classe « c ».

\bar{X}_c : matrice moyenne de la classe « c ».

G_i : matrice de covariance.

I. Démonstration que l'ACP2D est l'ACP appliquée sur les lignes de toutes les matrices de visages

$$\begin{aligned}
 G_i &= \frac{1}{M} \sum_{j=1}^M (X_j - \bar{X})^T (X_j - \bar{X}) \\
 &= \frac{1}{M} \left[(X_1 - \bar{X})^T (X_1 - \bar{X}) + (X_2 - \bar{X})^T (X_2 - \bar{X}) + \dots + (X_M - \bar{X})^T (X_M - \bar{X}) \right] \\
 &= \frac{1}{M} \left[(X_1 - \bar{X})^T (X_2 - \bar{X})^T \dots (X_M - \bar{X})^T \right] \left[(X_1 - \bar{X})^T (X_2 - \bar{X})^T \dots (X_M - \bar{X})^T \right]^T \\
 &= \frac{1}{M} \left[\hat{X}_1^T \hat{X}_2^T \dots \hat{X}_M^T \right] \left[\hat{X}_1^T \hat{X}_2^T \dots \hat{X}_M^T \right]^T \text{ avec } \hat{X}_j^T = X_j - \bar{X} \text{ pour } j=1, \dots, M \\
 &= \frac{1}{M} \left[\left[(\hat{X}_1^1)^T (\hat{X}_1^2)^T \dots (\hat{X}_1^n)^T \right] \dots \left[(\hat{X}_M^1)^T (\hat{X}_M^2)^T \dots (\hat{X}_M^n)^T \right] \right] \\
 &\quad \left[\left[(\hat{X}_1^1)^T (\hat{X}_1^2)^T \dots (\hat{X}_1^n)^T \right] \dots \left[(\hat{X}_M^1)^T (\hat{X}_M^2)^T \dots (\hat{X}_M^n)^T \right] \right]^T
 \end{aligned}$$

$$G_i = \frac{1}{M} \sum_{j=1}^M \sum_{i=1}^n (\hat{X}_j^i)^T (\hat{X}_j^i) \text{ avec } \hat{X}_j^i \text{ la } i^{\text{ème}} \text{ ligne de la } j^{\text{ème}} \text{ matrice centrée}$$

$$G_i = \frac{1}{M} \Psi^T \Psi$$

$$\text{Où } \Psi = \left[\left[(\hat{X}_1^1)^T (\hat{X}_1^2)^T \dots (\hat{X}_1^n)^T \right] \dots \left[(\hat{X}_M^1)^T (\hat{X}_M^2)^T \dots (\hat{X}_M^n)^T \right] \right]^T$$

Donc, la matrice G_i est la matrice de covariance de toutes les lignes des matrices de visages centrées de la base d'apprentissage.

II. Démonstration que l'ADL2D est l'ADL appliquée sur les colonnes de toutes les matrices de visages

1. pour la matrice de covariance inter-classes S_B

$$S_B = \sum_{c=1}^C n_c (\bar{X}_c - \bar{X})(\bar{X}_c - \bar{X})^T$$

$$S_B = n_1 (\bar{X}_1 - \bar{X})(\bar{X}_1 - \bar{X})^T + n_2 (\bar{X}_2 - \bar{X})(\bar{X}_2 - \bar{X})^T + \dots + n_C (\bar{X}_C - \bar{X})(\bar{X}_C - \bar{X})^T$$

Sans perte de généralité, on pose : $n_1 = n_2 = \dots = n_C = N$

$$S_B = N [(\bar{X}_1 - \bar{X}), (\bar{X}_2 - \bar{X}), \dots, (\bar{X}_C - \bar{X})][(\bar{X}_1 - \bar{X}), (\bar{X}_2 - \bar{X}), \dots, (\bar{X}_C - \bar{X})]^T$$

$$S_B = N \Phi_{S_B} \Phi_{S_B}^T$$

$$\text{avec } \Phi_{S_B} = [(\bar{X}_1 - \bar{X}), (\bar{X}_2 - \bar{X}), \dots, (\bar{X}_C - \bar{X})] \\ = [(\bar{X}_1(:,1) - \bar{X}(:,1)), \dots, (\bar{X}_1(:,m) - \bar{X}(:,m)) : \dots : (\bar{X}_C(:,1) - \bar{X}(:,1)), \dots, (\bar{X}_C(:,m) - \bar{X}(:,m))]$$

Où $\bar{X}_c(:,k)$ désigne la $k^{\text{ème}}$ colonne de la matrice moyenne de la classe « c » pour $c = 1, \dots, C$
 $\bar{X}(:,k)$: la $k^{\text{ème}}$ colonne de la matrice moyenne totale.

La matrice Φ_{S_B} est donc formée de la différence entre toutes ces colonnes.

2. pour la matrice de covariance intra-classe S_W

$$S_W = \sum_{c=1}^C \sum_{i=1}^{n_c} (X_i - \bar{X}_c)(X_i - \bar{X}_c)^T$$

$$S_W = [(X_1^1 - \bar{X}_1)(X_1^1 - \bar{X}_1)^T + (X_2^1 - \bar{X}_1)(X_2^1 - \bar{X}_1)^T + \dots + (X_{n_1}^1 - \bar{X}_1)(X_{n_1}^1 - \bar{X}_1)^T + \dots + \\ (X_1^C - \bar{X}_C)(X_1^C - \bar{X}_C)^T + (X_2^C - \bar{X}_C)(X_2^C - \bar{X}_C)^T + \dots + (X_{n_C}^C - \bar{X}_C)(X_{n_C}^C - \bar{X}_C)^T] \\ = [(X_1^1 - \bar{X}_1), (X_2^1 - \bar{X}_1), \dots, (X_{n_1}^1 - \bar{X}_1) : \dots : (X_1^C - \bar{X}_C), (X_2^C - \bar{X}_C), \dots, (X_{n_C}^C - \bar{X}_C)] \\ [(X_1^1 - \bar{X}_1), (X_2^1 - \bar{X}_1), \dots, (X_{n_1}^1 - \bar{X}_1) : \dots : (X_1^C - \bar{X}_C), (X_2^C - \bar{X}_C), \dots, (X_{n_C}^C - \bar{X}_C)]^T \\ = [\phi_1 \phi_2 \dots \phi_C][\phi_1 \phi_2 \dots \phi_C]^T$$

$$S_W = \Phi_{S_W} \Phi_{S_W}^T$$

avec $\Phi_{S_W} = [\phi_1 \phi_2 \dots \phi_C]$ et

$$\Phi_c = [(X_1^c(:,1) - \bar{X}_c(:,1)), \dots, (X_1^c(:,m) - \bar{X}_c(:,m)) : \dots : (X_{n_c}^c(:,1) - \bar{X}_c(:,1)), \dots, (X_{n_c}^c(:,m) - \bar{X}_c(:,m))]$$

où $X_i^c(:,k)$ désigne la $k^{\text{ème}}$ colonne de la $i^{\text{ème}}$ matrice de la classe « c » pour $c = 1, \dots, C$

$\bar{X}_c(:,k)$ la $k^{\text{ème}}$ colonne de la matrice moyenne la classe « c »

et par conséquent, la matrice Φ_{S_W} est la concaténation de la différence entre toutes les colonnes des images de visages.

Annexe D

Ce tableau résume des taux de reconnaissance obtenus par différentes méthodes de reconnaissance de visages appliquées sur plusieurs bases de visages:

Méthodes	Base de visages	Taux de Reconnaissance	Commentaires
Eigenfaces [51][50]	16 personnes	96.00 %	variations dans l'éclairage.
		85.00 %	variations dans la pose de la tête.
		64.00 %	variations de l'échelle du visage.
Eigenfaces [52]	AR:120 personnes, 13 images/personne	94.70 %	variations dans l'expression faciale.
		78.00 %	variations dans l'éclairage.
		66.20 %	variations dans le temps.
Eigenfaces [76]	Yale: 15 personnes, 11 images/personne	71.52 %	variations dans les conditions d'éclairage et les expressions faciales.
	ORL: 40 personnes, 10 images/personne	97.50 %	- variations dans la pose, l'éclairage, les expressions faciales et l'échelle du visage. - Leave-one-out.
Eigenfeatures [48]	45 personnes, 2 vues/personne	95.00 %	- visage neutre pour l'apprentissage, - visage souriant pour le test.
Modular eigenspaces [48]		98.00 %	
View-based eigenspace [48]	21 personnes 9 vues /personne	90.00 %	($\pm 90^\circ, 0^\circ \pm 45^\circ$): apprentissage ($\pm 68^\circ, \pm 23^\circ$): test
		83.00 %	($-90^\circ, 0^\circ -45^\circ$): apprentissage ($+68^\circ$ à 90°): test
Fisherfaces [57]	Yale: 16 personnes 10 images/personne	99.40 %	variations dans l'éclairage, les expressions faciales et les occultations.
Fisherfaces	ORL: 40 personnes 10 images/personne	94.50 % [52]	5 images/personne pour l'app.
		98.50 % [76]	Leave-one-out
Kernel eigenfaces	Yale: 15 personnes, 11 images/personne	94.00 % [52]	5 images/personne pour l'app.
		98.00 % [76]	Leave-one-out
Kernel eigenfaces [76]	Yale: 15 personnes, 11 images/personne	72.73 %	variations dans l'éclairage, les expressions faciales et les occultations.
ICA [76]		71.52 %	

ICA	ORL: 40 personnes 10 images/personne	85.50 % [52]	5 images/personne pour l'app.
		93.80 % [76]	Leave-one-out
2DPCA [52]	ORL: 40 personnes 10 images/personne	96.00 %	5 images/personne pour l'app.
		98.30%	Leave-one-out
	Yale: 15 personnes, 11 images/personne	84.24 %	- variations l'éclairage, les expressions faciales, - Leave-one-out
	AR: 120 personnes 13 images/personne	67.60 %	variations dans le temps.
		89.80 %	variations dans l'éclairage.
96.10 %		variations dans l'expression faciale.	
K2DPCA [83]	ORL: 40 personnes 10 images/personne	96.20 %	- 5 images/personne pour l'apprentissage. - variations dans la pose, l'éclairage, les expressions faciales et l'échelle du visage.
B2DPCA [83]		95.40 %	
U2DFDA [91]		96.30 %	
B2DFDA [91]		96.80 %	
KB2DFDA [91]		98.50 %	
2DLDA [77]	ORL: 40 personnes 10 images/personne	97.50 %	
2DLDA+LDA [77]		98.00 %	
2DLDA	PIE	99.32 %	6615 images de 63 personnes.
2DLDA+LDA		100.00 %	
HMM1D [103]	ORL: 40 personnes 10 images/personne	87.00 %	variations dans la pose, l'éclairage, les expressions faciales et l'échelle du visage.
HMM pseudo 2D [103]		95.00 %	
Approche géométrique [74]	47 personnes 4 images/personne	90.00 %	35 distances extraites automatiquement à partir des vues de face
Templates [74]		100.00 %	Modèles de visages.
Approche géométrique [89]	112 personnes	96.00 %	17 distances extraites à partir des vues de profil.
Mixture distances [65]	685 personnes	95.00 %	30 distances extraites manuellement.
Edge Map [49]	AR : 112 personnes 61 hommes/ 51 femmes	88.40 %	variations dans l'éclairage, les expressions faciales et l'échelle du visage.
Line Edge Map [49]		96.40 %	
Eigenfaces [49]		55.40 %	
Edge Map [49]	Bern : 30 personnes 10 images/personne	96.70 %	variations dans la pose de la tête : vues de faces, de profil gauche, de profil droite,...
Line Edge Map [49]		100.00 %	
Eigenfaces [49]		100.00 %	
WISARD 105]	16 personnes	100.00 %	système connexionniste.
Système de Baron [104]	42 personnes	100.00 %	visages orientés.

Tableau D.1. Comparaison entre des taux de reconnaissance obtenus par différentes méthodes de reconnaissance de visages.

Remarque :

K2DPCA: Kernel Two-dimensional Principal Component Analysis.

KB2DFDA: Kernel Bilateral Two-dimensional Fisher Discriminant Analysis.

Bibliographie

- [1] A.K.Jain, R.W.Duin and J.Mao. Statistical Pattern Recognition: A Rreview. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No.1, January 2000.
- [2] R.A. Fisher, Sc.D. F.R.S. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7:179-188, (1936).
- [3] Maurice Milgram. Reconnaissance des Formes. Méthodes Numériques et Connexionnistes. *Armand Colin edition*, 1993.
- [4] Sergios Theodoridis and Konstantions Koutroumbas. Pattern Recognition. Second Edition. *Elsevier Academic press*, USA, 2003.
- [5] Olivier Teyland. Apprentissage, Réseaux de Neurones et Applications. *Ph.D, université Lyon 2*, Décembre 2001.
- [6] Jack-Gerard Postaire. De l'image à la Décision. Analyse des Images Numériques et Théorie de la Décision. *Dunod*, Paris, 87.
- [7] D.Brylink and V.Starovoitov. Application of Recirculation Neural Network and Principal Component Analysis for Face Recognition. *The 2nd International Conference on Neural Networks and Artificial Intelligence*, pp.136-142, Minsk, Belarus, October 2-5, 2001.
- [8] Christian Gagné. Algorithmes Evolutionnaires Appliqués à la Reconnaissance de Formes et à la Conception Optique. *Ph.D, université Laval, Québec*, Mais 2005.
- [9] E.Fix and J.Hodges. Discriminatory Analysis Non Parametric Discrimination: Consistency Properties. *USAF School of Aviation Medecine. Project 21-49-004, Rep.4, pp. 261-279, 1951.*
- [10] Joaquim P. Marques de Sa. Pattern Recognition. Concepts, Methods and Applications. *Oporto University, Portugal, Springer*, May, 2001.
- [11] T. M. Cover, P. E. Hart. Nearest Neighbor Pattern Classification. *IEEE Transactions On Information Theory*, Vol. IT-IS, No. 1, January 1967.
- [12] M.M.Loève. Probability Theory. *Van Nostrand, Princeton*, 1955.
- [13] S.E.Fahlman. An Empirical Sudy of Learning Speed in Back-propagation Networks. *Technical report, CMU-CS-88-162*, September 1988.
- [14] L.B.Almeida. Supervised Models. Multilayer Perceptrons. *Handbook of Neural Computation. 1997 IOP Publishing Ltd and Oxford University Press.*
- [15] J.Sima. Introduction to Neural Networks. *Technical Report. N. V-755*, August 10, 1998.
- [16] Georg Thimm. Optimization of High Order Perceptrons. *Thesis of the Swiss Federal Institute of Technology (EPFL)*, 1996.

-
- [17] M.F.Redondo, C.H.Espinosa. Weight Initialisation Methods for Multilayer Feedforward. *ESANN'2001 Proceedings - European Symposium on Artificial Neural Networks Bruges (Belgium)*, 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 119-124.
- [18] W. Schiffmann, M. Joost, R. Werner. Optimization of the Backpropagation Algorithm for Training Multilayer Perceptrons. *Technical Report*, September 29, 1994
- [19] G. D. Magoulas, M. N. Vrahatis, G. S. Androulakis. Improving the Convergence of the Backpropagation Algorithm Using Learning Rate Adaptation Methods. *Neural Computation* 11, 1769–1796, 1999.
- [20] Rolf Pfeifer. *Neural Networks*, Chapter 4: Multilayer Perceptrons and Backpropagation.
- [21] M.Riedmiller, H.Braun. A Direct Adaptive Method for Faster Backpropagation Learning: the RPROP Algorithm. *In Proceedings of the IEEE International Conference on Neural Networks (ICNN)* (Ruspini, H. ed), pp.586-591, San Francisco, 1993.
- [22] Z. Zainuddin, N. Mahat, and Y. Abu Hassan. Improving the Convergence of the Backpropagation Algorithm Using Local Adaptive Techniques. *International Journal of Computational Intelligence*. Vol.1, No.3. 2004:1304-4508.
- [23] X.Lu and A.K.Jain. Automatic Feature Extraction for Multiview 3D Face Recognition. *To appear in Proc. International Conference on Automatic Face and Gesture Recognition*, Southsampton, UK, April 2006.
- [24] P.Coulibaly, F.Anctil et B.Bobée. Préviation Hydraulique par Réseaux de Neurones Artificiels : Etat de l'Art. *Cun.J.Civ.Eng.*26 .293-304, 1999.
- [25] Hervé Abdi. Neural Network Primer. *Journal of Biological System*, 2(3), 247-283, 1991.
- [26] F.Rosenblatt. The Perceptron: A Probabilistic Model for Information Storage and Retrieval in the Brain. *Psycho.Review*.Vol.65, 386-408, 1958.
- [27] P.Camon. Classification Supervisée par Réseaux de Neurones Multicouches. *Traitement du signal*. Vol.8, No.6, pp.387-407, 1991.
- [28] Eric Davalo et Patrick Naim. Des réseaux de Neurones. *Eyrolles 2^{ème} edition*, 1993.
- [29] B. Widrow and M.E.Hoff. Adaptive Switching Circuits. *In IRE WESCON Convention Record*, Volume 4, pages 96-104.IRE, New York, 1960.
- [30] W.S.McCulloch and W.Pitts. A logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5:115-133, 1943.
- [31] D.O.Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Wiley, New York, 1949.
- [32] M.L.Minsky and S.A. Papert. *Perceptrons*. MIT Press. Cambridge, MA, 1969.
- [33] R.A Jacobs. Increased Rates of Convergence Through Learning Rate Adaptation. *Neural Networks*, Vol.1, No.4, 295-307, 1988.
- [34] D.Plaut, S. Nowlan and G.Hinton. Experiments on Learning by Back Propagation. *Technical Report CMU-CS-86-126, Departement of Computer Science, Carnegie Mellon University*, Pittsburgh, Pennsylvania, 1986.
- [35] Y.Moses, Y.Adini and S.Ullman. Face Recognition: The Problem of Compensating for Changes in Illumination Direction. *European Conf.Computer Vision*, pp.286-296, 1994.
- [36] H.A.Rowley, S.Baluja and T.Kanade. Neural Network-Based Face Detection. Appears in *Computer Vision and Pattern Recognition*, 1996.
- [37] H.A.Rowley, S.Baluja and T.Kanade. Rotation Invariant Neural Network-Based Face Detection. *Computer Vision and Pattern Recognition*, pp.38-44, 1998.
- [38] Nicolas Eveno. Segmentation des Lèvres par un Modèle Déformable Analytique. *Ph.D, Institut National Polytechnique de Grenoble*, 2003.

- [39] V.Bao, R.Blouet. Vérification de l'Identité par les Données Biométriques. *Intl. Conf. RIVF'04*. February 2-5, Hanoi, Vietnam, 2004.
- [40] M.Marzouki. Enjeux des Techniques de Biométrie-Une 1^{ère} approche. *CNIL*, Septembre 2001.
- [41] M.Chassé. La Biométrie au Québec : les Enjeux. Document d'analyse. *Commission d'accès à l'information du Québec*. Juillet 2002.
- [42] R.P.Wildes. Iris Recognition: An Emerging Biometric Technology. *Proceedings of the IEEE*, Vol.85, No.9, Septembre 1997.
- [43] F.Perronnin et J-L.Duglay. Introduction à la Biométrie. Authentification des Individus par Traitement Audio-Vidéo. *Revue traitement du signal*. Vol.19, No.4. 2002.
- [44] A.Ross, A.Jain. Information Fusion in Biometrics. *Pattern Recognition Letters* 24(2003), 2115-2125.
- [45] A.Kumar, D.C. M. Wong, H.C. Shen, A.K. Jain. Personal Verification Using Palm Print and Hand Geometry Biometric. *In Proceedings of Fourth International Conference on AVBPA, (Guildford, U.K.)*, pp.668-678, June 2002.
- [46] B.Leroy, A.Chouakria, I.L.Herlin et E.Diday. Approche Géométrique et Classification pour la Reconnaissance de Visage, *INRIA*, France, 1996.
- [47] S.Z.Li and J.Lu. Face Recognition Using the Nearest Feature Line Method. *IEEE Transactions On Neural Networks*, 10(2), March, 1999, pp. 439-443.
- [48] A.Pentland, B.Moghaddam and T.Starner. View Based and Modular Eigenspaces for Face Recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, 1994.
- [49] Y.Gao, M.K.H Leung. Face Recognition Using Line Edge Map. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, Vol.24, No.6, June 2002.
- [50] M.A.Turk and A.P.Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71-86, 1991.
- [51] M.A.Turk and A.P.Pentland. Face Recognition Using Eigenfaces. *In Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 586-590, 1991.
- [52] J.Yang, D.Zhang, A.F.Frangi and J-Y.Yang. Two Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.26, No.1, January 2004.
- [53] G. Shakhnarovitch and B.Moghaddam. Face Recognition in Subspaces. *Hand Book of Face Recognition*, Springer-Verlag, 2004.
- [54] M.Visani, C.Garcia and J.M.Jolion. Two Dimensional-Oriented Discriminant Analysis for Face Recognition. *In Proc. of the Int. Conf. On Computer Vision and Graphics ICCVG'04* à paraître dans la série Computational Imaging and Vision, Varsovie, Pologne, Septembre 2004.
- [55] M.Visani, C.Garcia and J.M.Jolion. Face Recognition using Modular Bilinear Discriminant Analysis. *Dans Proceedings of the International Conference on Visual Information Systems (VIS 2005)*, pp.24-34, Amsterdam, Pays-Bas, Juillet 2005.
- [56] X.He, S.Yan, Y.Hu, P.Niyogi and H-J. Zhang. Face Recognition using Laplacianfaces. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol.27, No.3, March 2005.
- [57] P.N. Belhumeur, J.P. Hespanha and D.J.Kriegman. Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection. *In IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19, No.7, pages 711-720, July 1997.
- [58] W.Zhao, R.Chellappa and A.Krishnaswamy. Discriminant Analysis of Principal Components for Face Recognition. *In Proceedings, International Conference on Automatic Face and Gesture Recognition*.336-341, 1998.

- [59] W.Zhao, R.Chellappa and P.J.Philips. Subspace Linear Discriminant and Analysis for Face Recognition. *Tech.Rep.CAR-TR-914, Center for Automation Research, University of Maryland College Park, MD*, 1999.
- [60] W.Zhao, R.Chellappa, P.J.Philips and A.Rosenfeld. Face Recognition: A literature Survey. *ACM Computing Surveys*, Vol.35, No.4, pp.399-458, December 2003.
- [61] B.Moghaddam. Principal Manifolds and Bayesian Subspaces for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):780-788, June 2002.
- [62] V.Starovoitov, D.Samal, G.Votsis and S.Kollias. Face Recognition by Geometric Features. *Proceedings of 5th Pattern Recognition and Information Analysis Conference*, Minsk, May 1999.
- [63] H.Gu, G.Su and C.Du. Features Points Extraction from Faces. *Image and Vision Computing NZ*, Palmerston North, November 2003.
- [64] Z. Hammal, N. Eveno, A.Caplier, et P-Y.Coulon. Extraction Réaliste des Traits Caractéristiques du Visage à l'aide de Modèles Paramétriques Adaptés. *Colloque GRETSI sur le Traitement du Signal et des Images (GRETSI'03)*, Paris, France, Septembre 2003.
- [65] I.J.Cox, J.Ghosn and P.N.Yianilos. Feature-Based Recognition Using Mixture Distance. *In Proceedings, IEEE Conference on Computer Vision and pattern Recognition*. 209-216, 1996.
- [66] D.L.Swets and J.Weng. Using Discriminant Eigenfeatures for Image Retrieval. *IEEE Trans. on PAMI*, Vol.18, pp.831-836, Aug.1996.
- [67] N.Eveno, P.Delmas et P-Y.Coulon. Vers l'Extraction Automatique des Lèvres d'un Visage Parlant. *Colloque GRETSI sur le Traitement du Signal et des Images (GRETSI'01)*, Toulouse, France, Septembre 2001.
- [68] S.Sudi. Dimensionality Reduction Techniques: Study and Applications. *ECEG 313. Pattern Recognition*. Springer, 2005.
- [69] Abdelhalim Boualleg. La Reconnaissance Automatique des Visages. *Mémoire de magister*, université de Guelma, 2004.
- [70] S.Engelen, M.Hubert and K.V.Brandem. A Comparison of Three Procedures of Robust PCA in High Dimensions. *Australian Journal of Statistics*. Volume 34.No.2, 117-126, 2005.
- [71] M.E.Tipping and C.M.Bishop. Probabilistic PCA. *Microsoft Research*, September 27, 1999.
- [72] B.A.Drapper, K.Baek, M.S.Bartlett, J.R.Beveridge. Recognizing Faces with PCA and ICA. *Computer Vision and Image Understanding*, 91(1-2):115-137, July/Aug 2003.
- [73] M-H.Yang, D.J.Kriegman and N.Ahuja. Detecting Faces in Images: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24, No.1, January 2002.
- [74] R.Brunelli and T.Poggio. Face Recognition: Features Versus Templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol.15, pp.1042-1052, 1993.
- [75] S.Marcel and S.Bengio. Improving Face Verification using Skin Color Information. *In Proceeding of 16th International Conference On Pattern Recognition*. IEEE Computer Society Press, 2002.
- [76] M.H.Yang. Kernel Eigenfaces vs.Kernel Fisherfaces: Face Recognition Using Kernel Methods. *Proc. Fifth IEEE Int'l Conf. Automatic Face and Gesture Recognition (RGR'02)*, pp.215-220, May 2002.
- [77] J.Ye, R.Janardan, Q.Li. Two-Dimensional Linear Discriminant Analysis. *Advances in Neural Information Processing Systems (NIPS 2004)*, pp.1569-1576, 2004.
- [78] I.Craw, H.Ellis and J.Lishman. Automatic Extraction of face Features. *Pattern Recognition Letters*, Vol.5, pp.183-187, 1987.

- [79] T-P.Nguyen et T-H-L.Nguyen. Une Solution de Vérification de Visage basée sur l'Extraction des Caractéristiques de Type 'pic' et 'crête' de l'Image. *Intl. Conf. RIVF'04*, February 2-5, Hanoi, Vietnam, 2004.
- [80] N.Funabiki, M.Isogai, T.Higashino and M.Oda. An Eye Contour Extraction Algorithm From Face Image Using Deformable Template Matching. *Memoirs of the Faculty of Engineering Okayama university*, Vol.40, pp.78-82, January, 2006.
- [81] K.Peng, L.Cheng. A Robust Algorithm for Eye Detection On Gray Intensity Face Without Spectacles. *JCS &T*.Vol.5, No.3, pp.127-132, October 2005.
- [82] M.A.O.Vasilescu and D.Terzopoulos. Multilinear Analysis of Image Eensembles: TensorFaces. *In Proceedings of European Conference on Computer Vision*, pages 447-460, Copenhagen, Denmark, May 2002.
- [83] H.Kong, X.Li, L.Wang, E.K.Teoh, J.G.Wang and R.Venkateswarlu. Generalized 2D Principal Component Analysis. *In Proceedings of IJCNN*, 2005.
- [84] M.S.Bartlett, H.M.Lades and T.Sejnowski, Independent Component Representation for Face Recognition. *In Proceedings, SPIE Symposium on Electronic Imaging: Science and Technology*.528-539, 1998.
- [85] C.Liu and H.Wechler. Evolutionary Pursuit and its Application to Face Recognition. *IEEE Trans.Patt.Anal.Mach.Intell.* 22, 570-582, 2000.
- [86] S.H.Lin, S.Y.Kung and L.J.Lin. Face Recognition/Detection by Probabilistic Decision-Based Neural Network. *IEEE Transactions Neural Network*.8, 114-132, 1997.
- [87] H.Kong, L.Wang, E.K.Teoh, J-G.Wang and R.Venkateswarlu. A Frame Work of 2D Fisher Discriminant Analysis: Application to Face Recognition With Small Number of Training Samples. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pp.1083-1088, San Diego, USA, 20-25 June 2005.
- [88] A.J.Goldstein, L.D. Harmon and A.B.Lesk. Identification of Human Faces. *Proc.of the IEEE*, 59(5):748-760, 1971.
- [89] L.D.Harmon, M.K.Khan, R.Lasch and P.F. Raming. Machine Identification of Human Faces. *Pattern Recognition*, 13(2):97-110, 1981.
- [90] K.Sung and T.Poggio. Example-Based Learning for View-Based Human Face Detecction. *IEEE Trans.Patt Anal. Mach. Intell.*, 20, 39-51, 97.
- [91] H.Kong, E.K.Teoh, J-G.Wang and C.Kambhamettu. Generalized 2D Fisher Discriminant Analysis. British Machine Vision Conference, BMVC, September 2005, Oxford, UR.
- [92] D.J.Beymer. Face Recognition Under Varying Pose. *Tech.Report.1461.MIT AI Lab, Massachusetts Institute of Technology*, Cambridge, MA, 1993.
- [93] L.Wiscott, J-M.Fellous and V.DerMalsburg. Face Recognition by Elastic Bunch Graph Matching. *IEEE Trans.Patt Anal. Mach. Intell.*, 19,775-779, 1997.
- [94] K-W.Wong, K-M.Lam, W-C.Siu, An Efficient Algorithm for Human Face Detection and Facial Feature Extraction under Different Conditions. *Centre for Multimedia Signal Processing, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, Pattern Recognition* 34, 2001.
- [95] W.Zhao and R.Chellappa. SFS Based View Synthesis for Robust Face Recognition. *In Proceedings International Conference on Automatic Face and Gesture Recognition*, 2000.
- [96] P.J.Philips, H.Moon, S.Rizvi and P.Rauss. The Feret Evaluation Methodology for Face Recognition Algorithms. *IEEE Trans.Patt Anal. Mach. Intell.*, 22, 2000.
- [97] T.Cootes, C.Taylor, D.Cooper and J.Graham. Active Shape Models-Their Ttraining and Application. *Comput.Vis.Image Understand*.61, 18-23, 1995.
- [98] T.F.Cootes, G.J.Edwards and C.J.Taylor. Active Appearance Models. *Proc. European Conference on Computer Vision*.Vol.2, pp.484-498, Springer, 98.

-
- [99] A.Lanitis, C.J.Taylor and T.F Cootes. Automatic Face Identification System Using Flexible Appearance Models. *Image Vis. Comput.*13, 393-401, 1995.
- [100] P.W.Hallinan. Reconizing Human Eyes. In SPIE Proceedings, Vol.1570: Geometric Methods. *In Computer Vision.*214-226, 1991.
- [101] A.L Yuille, D.S Cohen and P.W.Hallinan. Feature Extraction from Faces Using Deformable Templates. *Int.J.Comput.Vis.*8, 99-112, 1992.
- [102] T.Kanade. Picture Processing by Computer Complex and Recognition of Human Faces. *Tech. Report. Kyoto University, Dep.of Information Science*, 1973.
- [103] Ferdinando Silvestro Samaria. Face Recognition using Hidden Markov Models. *Ph.D. Ddissertation. University of Cambridge*, Cambridge, U.K, 1994.
- [104] R.J.Baron. Mechanisms of Human Facial Recognition. *International Journal on Man Machine Studies*, 15:137-178, 1981.
- [105] T.J.Stonham. Practical Face Recognition and Verification with WISARD. *In H.D.Ellis, M.A.Jeeves, F.Newcombe, and A.Young, editors, Aspects of face processing*, pages 426-441. Martinus Nijhoff Publishers, 1986.
- [106] S. Raudys and R.P.W. Duin. On Expected Classification Error of the Fisher Linear Classifier with Pseudo-Inverse Covariance Matrix. *Pattern Recognition Letter*, 1998.
- [107] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. Regularized Discriminant Analysis for the Small Sample Size Problem in Face Recognition. *Pattern Recognition Letter*, 2003.
- [108] P. Howland and H. Park. Generalizing Discriminant Analysis using the Generalized Singular Value Decomposition. *IEEE Trans. on PAMI*, 2004.
- [109] H.Yu and J.Yang. A Direct LDA Algorithm for High-Dimensional Data with Application to Face Recognition. *Pattern Recognition*, 2001.
- [110] R. Huang, Q.S. Liu, H.Q. Lu, and S.D. Ma. Solving the Small Sample Size Problem of LDA. *In Proceedings of ICPR*, 2002.
- [111] B.Moghaddam and A.Pentland. Probabilistic Visual Learning for Object Representation. *IEEE Trans. Patt. Anal.Mach. Intell*, 19, 696-710, 1997.

Résumé : En plus des variations d'illumination, de la pose de la tête, des expressions faciales, des occultations partielles, et du vieillissement,... les approches globales, utilisant le visage entier pour la reconnaissance, souffrent de la grande taille de données à traiter due à la transformation des images de visages en vecteurs. Pour corriger ce problème majeur, nous proposons une approche hybride bidimensionnelle **ACPDL2D**, intégrée dans un système automatique de reconnaissance faciale comportant trois étapes :**i)** localisation automatique de visages, **ii)** extraction de caractéristiques et réduction de dimensions par l'ACPDL2D et **iii)** une classification par un PMC. Le système proposé **Neuro-ACPDL2D** assure ainsi un codage et une compression de données très efficace, via une projection hybride bilatérale dans un espace bidimensionnel, dépassant celle des autres méthodes récemment proposées tout en garantissant une reconnaissance satisfaisante.

Dans le but de corriger le problème des occultations partielles, nous avons également proposé des **approches locales bidimensionnelles** avec une classification neuronale. Elles reposent sur la forme des éléments faciaux au sens où elles s'appliquent, séparément, sur deux régions faciales (une contenant les yeux avec les sourcils et l'autre le nez), après une localisation automatique de ces derniers. Ces approches sont plus performantes que les approches géométrique et globales bidimensionnelles en termes d'espace mémoire et de reconnaissance, et elles sont privilégiées car invariantes aux arrières plans, au style de coiffure, moins sensibles aux variations d'illumination et présentent l'avantage de pouvoir identifier une personne dont le visage est en partie cachée.

Mots clés : ACP2D, ADL2D, Approches locales bidimensionnelles, Localisation, Eléments faciaux, Neuro-ACPDL2D, PMC, Reconnaissance automatique de visages, Séquences vidéo.