Vers de nouvelles approches discriminantes pour la reconnaissance automatique de visages

Muriel Visani

N° d'ordre 2005-ISAL-0094 Année 2005

THÈSE

pour obtenir le grade de

Docteur de l'Institut National des Sciences Appliquées de Lyon

Spécialité Informatique

Ecole doctorale Informatique et Information pour la Société (EDIIS)

Vers de nouvelles approches discriminantes pour la reconnaissance automatique de visages

par

Muriel VISANI

Thèse soutenue le 25 Novembre 2005

Composition du jury

POGGI Jean-Michel	Professeur – Université de Paris 5	Président
JOLION Jean-Michel	Professeur – INSA de Lyon	Directeur
GARCIA Christophe	Chercheur – France Télécom R&D	Directeur
LALLICH Stéphane	Professeur – Université de Lyon 2	Rapporteur
POSTAIRE Jack-Gérard	Professeur – Université de Lille 1	Rapporteur
DUGELAY Jean-Luc	Professeur – Institut Eurécom	Examinateur

2005

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
	CHIMIE DE LYON Responsable : M. Denis SINOU	M. Denis SINOU Université Claude Bernard Lyon 1 Lab Synthèse Asymétrique UMR UCB/CNRS 5622 Bât 308 2ème étage 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cedex Tél : 04.72.44.81.83 Fax : 04 78 89 89 14 sinou@univ-lyon1.fr
E2MC	**ECONOMIE, ESPACE ET MODELISATION DES COMPORTEMENTS** Responsable : M. Alain BONNAFOUS	M. Alain BONNAFOUS Université Lyon 2 14 avenue Berthelot MRASH M. Alain BONNAFOUS Laboratoire d'Economie des Transports 69363 LYON Cedex 07 Tél : 04.78.69.72.76 Alain.bonnafous∂ish-lyon.cnrs.fr
E.E.A.	**ELECTRONIQUE, ELECTROTECHNIQUE, AUTOMATIQUE** M. Daniel BARBIER	M. Daniel BARBIER INSA DE LYON Laboratoire Physique de la Matière Bâtiment Blaise Pascal 69621 VILLEURBANNE Cedex Tél : 04.72.43.64.43 Fax 04 72 43 60 82 Daniel.Barbier@insa-lyon.fr
E2M2	**EVOLUTION, ECOSYSTEME, MICROBIOLOGIE, MODELISATION** **http://biomserv.univ-lyon1.fr/E2M2** **M. Jean-Pierre FLANDROIS**	M. Jean-Pierre FLANDROIS UMR 5558 Biométrie et Biologie Evolutive Equipe Dynamique des Populations Bactériennes Faculté de Médecine Lyon-Sud Laboratoire de Bactériologie BP 1269600 OULLINS Tél : 04.78.86.31.50 Fax 04 72 43 13 88 E2m2∂biomserv.univ-lyon1.fr
EDIIS	**INFORMATIQUE ET INFORMATION POUR LA SOCIETE** http://www.insa-lyon.fr/ediis M. Lionel BRUNIE	M. Lionel BRUNIE INSA DE LYON EDIIS Bâtiment Blaise Pascal 69621 VILLEURBANNE Cedex Tél : 04.72.43.60.55 Fax 04 72 43 60 71 ediis@insa-lyon.fr
EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTE** http://www.ibcp.fr/ediss M. Alain Jean COZZONE	M. Alain Jean COZZONE IBCP (UCBL1) 7 passage du Vercors 69367 LYON Cedex 07 Tél : 04.72.72.26.75 Fax : 04 72 72 26 01 cozzone@ibcp.fr
	MATERIAUX DE LYON http://www.ec-lyon.fr/sites/edml M. Jacques JOSEPH	M. Jacques JOSEPH Ecole Centrale de Lyon Bât F7 Lab. Sciences et Techniques des Matériaux et des Surfaces 36 Avenue Guy de Collongue BP 163 69131 ECULLY Cedex Tél : 04.72.18.62.51 Fax 04 72 18 60 90 Jacques.Joseph@ec-lyon.fr
Math IF	**MATHEMATIQUES ET INFORMATIQUE FONDAMENTALE** http://www.ens-lyon.fr/MathIS M. Franck WAGNER	M. Franck WAGNER Université Claude Bernard Lyon1 Institut Girard Desargues UMR 5028 MATHEMATIQUES Bâtiment Doyen Jean Braconnier Bureau 101 Bis, 1er étage 69622 VILLEURBANNE Cedex Tél : 04.72.43.27.86 Fax : 04 72 43 16 87 wagner@desargues.univ-lyon1.fr
MEGA	**MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE** http://www.lmfa.ec-lyon.fr/autres/MEGA/index.html M. François SIDOROFF	M. François SIDOROFF Ecole Centrale de Lyon Lab. Tribologie et Dynamique des Systêmes Bât G8 36 avenue Guy de Collongue BP 163 69131 ECULLY Cedex Tél :04.72.18.62.14 Fax : 04 72 18 65 37 Francois.Sidoroff@ec-lyon.fr

Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude à mes directeurs de thèse Christophe Garcia, ingénieur de recherche et expert pour le groupe France Télécom R&D et Jean-Michel Jolion, professeur à l'INSA de Lyon, pour m'avoir encadrée et guidée avec clairvoyance, pour leurs nombreux conseils, leur confiance et leur soutien constant tout au long de ces trois années de thèse. Je les remercie également pour leur disponibilité, leurs très grandes qualités humaines et leur relecture minutieuse du manuscrit. Merci.

Je remercie vivement Jean-Michel Poggi, professeur à l'Université de Paris 5, pour avoir accepté de présider mon jury de thèse. Je suis également très reconnaissante à Stéphane Lallich, professeur à l'Université de Lyon 2 et à Jack-Gérard Postaire, professeur à l'Université de Lille 1, de s'être penchés avec rigueur et grand intérêt sur ce travail et de m'avoir fait l'honneur d'en être les rapporteurs. Mes plus sincères remerciements vont également à Jean-Luc Dugelay, professeur à l'Institut Eurécom (Sophia-Antipolis), qui a examiné ce travail de thèse et a participé au jury.

Ce travail de thèse a été réalisé dans le cadre d'un CDD de formation par la recherche France Télécom R&D, que je remercie pour m'avoir fourni d'excellentes conditions logistiques et financières. Lors de ces trois années de thèse, j'ai évolué au sein des laboratoires LIRIS (Unité Mixte de Recherche 5205) et IRIS (France Télécom R&D, Rennes). Je remercie leurs directeurs respectifs Bernard Peroche et Vincent Marcatté ainsi que tous les chercheurs de ces laboratoires pour leur accueil chaleureux et pour m'avoir fait profiter de leur expérience et de leur savoir. Grâce à eux, j'ai eu la chance de travailler dans un environnement de recherche extrêmement convivial et motivant.

Je remercie également Henri Sanson, responsable de l'équipe CIM du laboratoire IRIS, pour l'intérêt constant qu'il a porté à mon travail.

Merci à Sid-Ahmed Berrani pour sa relecture détaillée des premiers chapitres du manuscrit. Merci également à toutes les personnes qui m'ont entourée et aidée durant ces trois années de travail et notamment à tous les membres de l'équipe CIM.

Un grand merci à mes parents, à ma sœur et à toute ma famille pour leur soutien indéfectible. Enfin, rien n'aurait été possible sans Eric, qui m'a épaulée et encouragée dans les instants difficiles et a su partager avec moi les moments de joie rencontrés lors de la préparation de cette thèse.

à
Ugo, Venturina et Marie
mes parents
Eric

Table des matières

Introduction générale

Contexte

Reconnaître un visage, c'est lui affecter une identité parmi celles d'un ensemble de visages connus. Les humains sont dotés d'une excellente aptitude à identifier leurs semblables. Les études biologiques tendent à prouver que la reconnaissance humaine des visages constitue un processus spécifique de reconnaissance d'objets, qui serait mené dans une région particulière du cerveau. On peut considérer qu'il en est de même de la reconnaissance automatique, qui constitue un domaine particulier du traitement d'images et de la reconnaissance de formes. Ses spécificités proviennent surtout de la nature des objets à différencier. En effet, les visages de deux personnes différentes sont structurellement très proches, car dotés des mêmes caractéristiques faciales (yeux, nez, bouche), dont la localisation varie très peu. De plus, les sources de variabilité entre deux vues d'un même visage sont multiples, et peuvent même engendrer des dissimilarités plus importantes que celles observées entre deux visages différents. Aussi peut-on considérer qu'il s'agit d'une tâche de classification plus complexe que la reconnaissance d'objets génériques. En effet, cette dernière consiste généralement à classer un objet observé dans sa catégorie d'appartenance. Dans le cas des visages, il s'agirait de classer un visage dans la catégorie des visages, tâche que nous désignerons dans la suite par le terme *détection de visages*. En revanche, dans le cadre de la reconnaissance, nous connaissons la nature de l'objet mais cherchons à le mettre en correspondance avec les objets de sa catégorie qui lui sont le plus similaires. Il existe donc une différence fondamentale entre la reconnaissance d'objets génériques et la reconnaissance de visages qui nécessite une classification à un niveau supérieur.

Motivation

Vingt années de recherche intensive dans le domaine de la reconnaissance automatique de visages dans des images numériques ont abouti à la conception de systèmes d'*authentification* performants. Dans ce contexte, la personne qui se présente au système prétend avoir une certaine identité et le processus doit être capable de déterminer de manière fiable si l'identité revendiquée est ou non authentique. Par contre, dans un cadre d'*identification* plus générale, où l'on ne dispose d'aucune information *a priori* concernant l'identité du visage, la plupart des systèmes connaissent une baisse de leurs performances dans des conditions réelles d'application. Nous nous intéresserons dans cette thèse à l'identification de visages humains, en monde *fermé* ou *ouvert*. Le monde est dit fermé si tout visage présenté au système est enregistré dans la base de connaissance, et ouvert sinon. Le nombre d'applications potentielles est très important. Il n'y avait pas à l'initiative de cette thèse d'application particulièrement visée. D'où la nécessité de concevoir un système qui soit le plus universel possible, c'est-à-dire qui ne dépende pas de l'ajustement d'un nombre important de paramètres en fonction des caractéristiques des bases de

visages utilisées (telles que la taille de la base par exemple).

Problématique

La reconnaissance automatique de visages nécessite à la fois la mise en œuvre de traitements des images (localisation/segmentation du visage, correction d'illumination, etc.) [BJL03] et de techniques d'apprentissage et de discrimination [Fuk90]. On considère que les traitements des images sont effectués en amont de la reconnaissance. Dans cette thèse, nous nous concentrerons plus particulièrement sur les phases de modélisation et de classification. La figure 1 schématise le processus de reconnaissance de visages qui, comme toute tâche de reconnaissance d'objets, se décompose en deux étapes :

- extraction d'éléments caractéristiques (*signatures*) ;
- classification des signatures. L'ensemble des images de visages connus est stocké dans une *base de connaissance*. Cette base de connaissance peut contenir plusieurs images d'une même personne, sous des conditions de prise de vue différentes. Chaque image est étiquetée par son identité associée. À chaque personne de la base de connaissance, on associe une *signature* qui lui est caractéristique. La reconnaissance d'un visage-requête se fait en deux étapes. Dans une première phase, on extrait sa signature à l'aide de la même technique que celle appliquée à la base d'apprentissage. Puis, la signature-requête ainsi obtenue est mise en correspondance avec la signature de la personne la plus proche dans la base de connaissance. On en déduit l'identité du visage-requête.

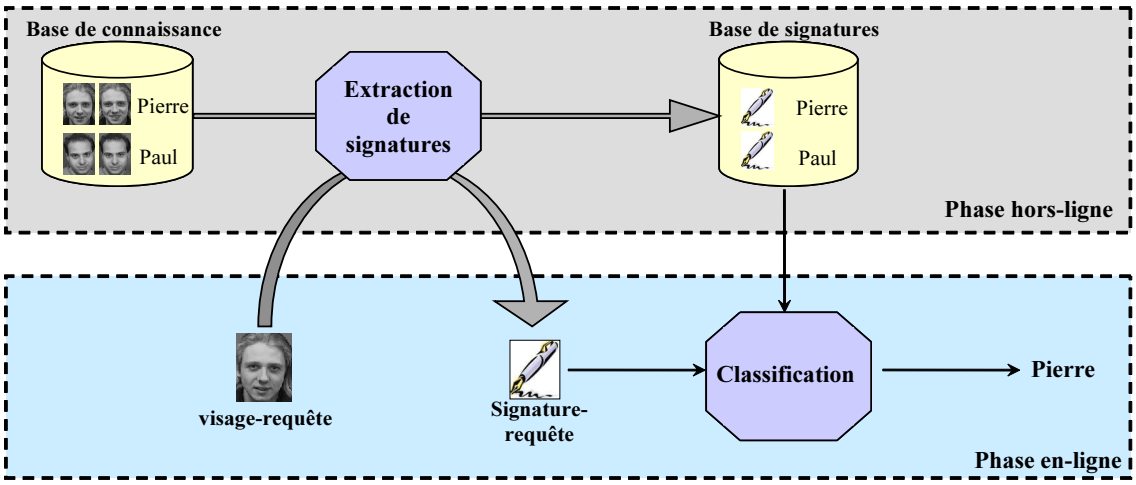


FIG. 1 – *Schématisation du processus de reconnaissance de visages.*

Pour le choix des signatures, on privilégie un certain nombre de propriétés. Celles-ci doivent être *invariantes* et *discriminantes*. Le terme *invariant* s'applique à des caractéristiques qui ne sont pas ou peu affectées par des changements acceptables[1] d'apparence. En d'autres termes, les signatures extraites de deux vues d'une même personne doivent être le plus proche possible (au sens du critère utilisé pour leur classification), même si les conditions de prise de vue sont

1. Le terme « acceptable » est volontairement vague ; en effet, le seuil d'acceptabilité des changements varie en fonction de la méthode choisie et de l'application visée.

différentes. Le terme *discriminant* indique que ces caractéristiques doivent, de plus, prendre des valeurs significativement différentes pour les vues de deux personnes distinctes.

Notons que le choix de la méthode de classification est très dépendant de la nature des signatures et de l'application visée et a globalement donné lieu à moins de recherches que l'extraction de celles-ci.

Contributions et plan du manuscrit

Les contributions de cette thèse concernent à la fois l'extraction de signatures caractéristiques des visages et la classification de celles-ci.

Tout d'abord, nous introduisons une relecture des très nombreux travaux relevant de ce domaine qui permettra, nous l'espérons, à tout chercheur d'appréhender la complexité de ce domaine et de ses avancées récentes. Concernant l'extraction de signatures, nous introduisons dans un premier temps une approche originale, basée sur une Analyse Discriminante Linéaire et une modélisation bidimensionnelle orientée (2Do) des données. Cette approche globale, baptisée Analyse Discriminante Linéaire Bidimensionnelle Orientée (ADL2Do), se décline en deux versions, selon que l'on considère les lignes ou les colonnes des images. La représentation 2Do permet de mettre en œuvre l'analyse de données directement sur les lignes ou les colonnes de l'image, et ainsi d'éviter implicitement le problème de la singularité inhérent à la sous-représentation des données de visages. Après avoir choisi une mesure de dissimilarité permettant d'obtenir de très bons résultats de classification, nous montrerons la complémentarité de ces deux techniques de reconnaissance, ce qui ouvre la voie à leur fusion. Nous introduirons dans une seconde étape une méthode nommée *Analyse Discriminante Bilinéaire* (ADB), qui constitue un mode de combinaison efficace des deux approches orientées. Il s'agit d'une technique globale bidimensionnelle, au sens où elle allie les avantages des approches orientées en ligne et en colonne. Après avoir sélectionné une mesure de dissimilarité adaptée, nous mettrons en évidence ses très bonnes performances pour l'identification en monde fermé et la comparerons aux techniques de l'état de l'art. Afin de garantir une tolérance accrue à des variations d'expression faciale, nous introduirons dans un troisième temps une approche hybride nommée *ADB Modulaire* (ADBM). Celle-ci est basée sur l'utilisation conjointe de trois experts ADB, construits indépendamment sur des régions faciales différentes. Plusieurs modes de combinaison des experts seront étudiés. Nous mettrons en évidence ses très bonnes performances pour l'identification de visages en monde fermé.

Notre contribution dans le contexte de la classification provient en premier lieu de l'étude de l'efficacité de différentes mesures de similarité pour la classification des signatures obtenues par ADL2Do et par ADB. Nous étudierons notamment les performances des distances de Minkowski et des mesures de Minkowski fractionnaires, et montrerons qu'une stratégie d'affectation au plus proche voisin est plus efficace qu'à la plus proche moyenne. Cette règle d'affectation présente cependant le désavantage d'être coûteuse, et potentiellement influencée par des observations aberrantes. Nous introduisons donc dans un second temps l'utilisation de Réseaux de Fonctions à Base Radiale Normalisés (RFBRN), qui permettent de modéliser les classes de signatures avec un faible nombre de paramètres, et ainsi de réduire la complexité en termes de temps de calcul de la phase de classification par rapport à un plus proche voisin. De plus, ce type de réseau de neurones permet de mieux prendre en compte les distributions des classes lors de la classification, ce qui engendre une robustesse accrue à d'éventuelles observations aberrantes. Cela nous permet également de définir des règles de décision simples mais efficaces qui nous permettront d'appliquer l'approche ainsi définie dans un contexte en monde ouvert. Une évaluation rigoureuse de l'approche proposée permettra de mettre en lumière ses très bonnes performances

en monde fermé comme ouvert et sa tolérance à un ensemble de sources de variations récurrentes dans le problème de la reconnaissance de visages (changements de pose, occultations partielles des visages, etc.)

Afin de détailler ces apports, nous avons choisi d'articuler notre étude autour de cinq chapitres principaux.

Le premier chapitre vise à positionner précisément le problème ainsi que les enjeux et les applications possibles de la reconnaissance de visages. Nous étudierons également le système de reconnaissance humain de visages, afin de tirer les meilleurs enseignements des excellentes aptitudes du cerveau humain dans ce domaine. Puis, nous présenterons les principaux protocoles d'évaluation des systèmes automatiques et nous mettrons en lumière le nombre important de difficultés inhérentes à la reconnaissance de visages.

Dans le deuxième chapitre, nous étudierons les principales techniques de l'état de l'art proposées pour la reconnaissance de visages humains. Nous présenterons les performances de ces techniques, et discuterons de leurs avantages et de leurs inconvénients.

Dans le troisième chapitre, nous nous focaliserons sur l'étude des approches basées sur l'Analyse Discriminante Linéaire ainsi que ses variantes les plus utilisées dans le contexte de la reconnaissance de visages.

Le quatrième chapitre vise dans un premier temps à mettre en évidence les avantages de la représentation 2Do des données tant en termes de performance que de tolérance à différentes sources de variations. Puis, nous présenterons les deux versions issues de l'Analyse Discriminante Linéaire Bidimensionnelle orientée (ADL2Do) et nous mettrons en évidence leurs très bonnes performances et leur complémentarité.

Dans le cinquième chapitre, nous introduirons la technique d'Analyse Discriminante Bilinéaire (ADB), puis montrerons qu'il s'agit d'un mode de fusion efficace des deux versions de l'ADL2Do. Par la suite, nous présenterons l'ADB Modulaire, et mettrons en lumière ses très bonnes performances et notamment sa tolérance accrue à des changements d'expression faciale. Nous montrerons également l'efficacité de l'utilisation de RFBRN pour la classification de signatures issues de l'ADB dans le contexte de l'identification de visages en monde ouvert.

Une conclusion terminera ce mémoire et introduira les principales perspectives de ce travail de recherche.

Chapitre 1

La reconnaissance automatique de visages : problématiques et applications

1.1 Introduction

Les premiers travaux concernant la reconnaissance automatique de visages remontent au début des années 1970. Les techniques introduites à l'époque utilisaient pour la plupart des mesures estimées autour des éléments faciaux des visages [Ble66, Kel70, Kan77]. Mais ce n'est qu'au début des années 1990 que le volume de recherche concernant la reconnaissance de visages a réellement commencé à croître. Par la suite, l'engouement pour cette problématique n'a fait qu'augmenter, si bien qu'aujourd'hui la reconnaissance de visages constitue l'un des domaines les plus explorés de la reconnaissance de formes et de l'analyse d'images. Cet intérêt croissant s'est soldé par l'apparition de conférences internationales sur le sujet, telles que l'*International Conference on Automatic Face and Gesture Recognition* (AFGR) en 1995 et l'*International Conference on Audio-and Video-Based Authentication* (AVBPA) en 1997. La figure 1.1 montre l'évolution dans le temps des publications référencées dans la base de recherche IEEE Xplore[2], et ayant trait à la reconnaissance de visages.

Cette tendance peut être expliquée par des enjeux croissants, notamment dans les domaines de l'indexation et de la sécurité, mais aussi par les avancées technologiques réalisées les trente dernières années dans le domaine de l'analyse d'images. Ces dernières permettent aujourd'hui de proposer des solutions exploitables –au moins partielles– à ce problème. Le domaine de la reconnaissance de visages continue à attirer de nombreux chercheurs issus de disciplines telles que le traitement d'images, la reconnaissance de formes, les réseaux de neurones, la vision par ordinateur, les interfaces homme/machine, la neurophysiologie et la neuropsychologie.

Malgré les efforts engagés, on ne peut pas considérer à ce jour que la reconnaissance automatique de visages soit un problème résolu, comme le montre la récente évaluation des principales techniques proposées dans ce contexte, menée par le National Institute of Standards and Technology (NIST) [PGM+03]. En effet, malgré la maturité des meilleures techniques évaluées, celles-ci peuvent être insuffisantes dans le contexte d'applications réelles, où les sources de variabilité sont multiples (illumination, angle de prise de vue, etc.). Étant donné que les visages sont des objets structurellement très proches, ces changements peuvent engendrer des différences plus importantes entre deux vues d'un même visage qu'entre deux vues de visages différents. Cela fait de la reconnaissance automatique de visages un problème de classification particulièrement difficile.

2. Cette base de recherche référence les articles parus dans les journaux, magazines et actes de conférence IEEE (et IEE) depuis 1988.

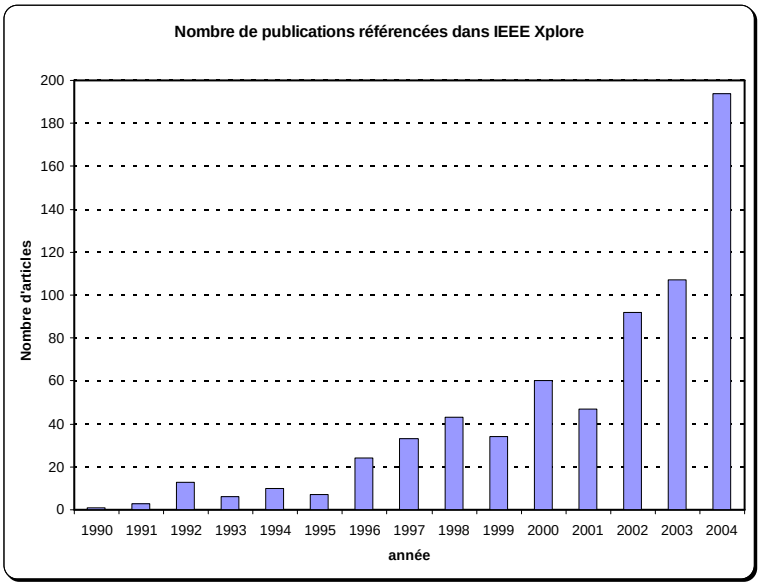


FIG. 1.1 – *Nombre de publications référencées par le moteur de recherche IEEE Xplore, en fonction de l'année de publication. Les titres des articles comptabilisés contiennent au moins l'un des termes-clés suivants : Face Recognition, Face Identification, Face Authentication, Face Representation et Face Verification. On note un engouement croissant depuis le début des années 1990, avec un accroissement exponentiel du nombre de publications en 2004.*

Ce chapitre est organisé comme suit. Après avoir précisément défini la problématique en section 1.2, nous présenterons en section 1.3 ses principales applications et les énormes enjeux associés. En section 1.4, nous nous pencherons sur le processus de reconnaissance humaine de visages, afin de tirer les meilleurs enseignements des excellentes aptitudes du cerveau humain dans ce domaine. La section 1.5 met en évidence les principales sources de difficultés inhérentes à la reconnaissance automatique. En section 1.6 nous présenterons les principaux prétraitements des images nécessaires à la mise en œuvre de la plupart des systèmes de reconnaissance. Enfin, les techniques d'évaluation de ces systèmes seront étudiées en section 1.7.

1.2 Position du problème

La plupart des algorithmes de reconnaissance automatique de visages portent sur la classification d'*images fixes* 2D. On dispose d'une *base de connaissance* contenant des photographies de personnes connues, c'est-à-dire les personnes que le système est censé reconnaître lors de toute apparition ultérieure. Lorsqu'un *visage-requête* (image d'un visage à reconnaître) est présenté au système, celui-ci va chercher à lui assigner une identité parmi celles contenues dans la base de connaissance. Le système de reconnaissance est basé sur un modèle généralement construit à partir d'une *base d'apprentissage* contenant un ensemble d'images de visages, par le biais d'un algorithme d'apprentissage. Dans certains cas, la base d'apprentissage et la base de connaissance sont confondues. Le modèle est alors spécifiquement conçu pour les visages qu'il vise à reconnaître. Dans d'autres cas au contraire, soit du fait des spécificités de la technique utilisée, soit à cause de l'application, les bases d'apprentissage et de connaissance sont distinctes. La plupart des systèmes sont basés sur l'utilisation des niveaux de gris. Les quelques algorithmes prenant en

compte la couleur ne sont pour la plupart que des généralisations à trois canaux des techniques utilisées en niveaux de gris [TRL99].

Ces dernières années, la reconnaissance de visages depuis des *vidéos* commence à être explorée [ZCPR03]. Un avantage de ces applications est que l'on peut, par le biais de l'utilisation d'un module de *suivi* des visages, disposer d'un nombre important de vues du visage-requête suivi. La plupart des systèmes reposent sur l'utilisation de techniques initialement conçues dans le contexte d'images fixes. La plupart du temps, on se ramène à un problème de classification d'images fixes par l'utilisation de l'une des trois stratégies suivantes. La première consiste à sélectionner une image-clé représentative de l'ensemble des vues du visage-requête. La deuxième solution est d'utiliser un algorithme de vote sur les résultats de classification obtenues pour chacune de ces vues [GMP00]. La troisième possibilité est de construire un modèle spécifique pour chaque personne de la base de connaissance, ainsi que pour le visage-requête, à partir de l'ensemble des vues de la personne considérée dont on dispose. Les caractéristiques du modèle du visage-requête sont alors comparées à celles des modèles de la base de connaissance afin de lui assigner l'identité du modèle le plus proche [YFM98, TB99a, TLV00, WS03].

Très récemment, les avancées dans le domaine de l'acquisition de données tridimensionnelles (notamment par le biais de scanners laser) ont permis l'émergence de technologies de *reconnaissance 3D*. Il existe deux grandes familles de méthodes : celles basées sur la mise en correspondance d'objets 3D directement [CBF05], et celles se ramenant pour la classification à un problème de reconnaissance 2D [RBBV04]. Le premier type de méthodes nécessite généralement que l'on dispose d'images 3D aussi bien pour les visages de la base d'apprentissage que pour les visages-requêtes, tandis que dans le second cas seules les vues 3D de la base d'apprentissage sont requises. Quelle que soit la solution retenue, il est nécessaire de collecter un certain nombre de visages par le biais de capteurs 3D, ce qui réduit le champ des applications. En effet, de tels capteurs reposent encore à ce jour sur une prise de vue *intrusive*, au sens où elle requiert la coopération du sujet.

Dans le cadre de cette thèse, nous nous intéressons aux techniques basées sur l'étude d'images fixes 2D des visages, caractérisées par des valeurs de pixels en niveaux de gris. La plupart des méthodes reportées, proposées et/ou évaluées dans le présent document sont conçues pour la reconnaissance de visages dans des images fixes, mais comme nous l'avons vu plus haut elles peuvent facilement être étendues à la reconnaissance dans des vidéos.

1.3 Applications et enjeux

L'impact stratégique et les enjeux financiers du marché mondial de l'information électronique sont énormes. Il est essentiel de sécuriser les transactions en protégeant l'identité des utilisateurs d'une part, et d'analyser le contenu informationnel d'autre part. Par ailleurs, la demande de sécurisation des lieux publics s'est faite de plus en plus pressante ces dernières années. Ces éléments expliquent en grande partie l'effort de recherche investi dans les techniques de reconnaissance de visages. Face à cette demande, de nombreuses entreprises ont proposé des produits commerciaux, dont les principaux sont répertoriés en table 1.1.

Dans la suite de cette section, nous présenterons les principales applications de la technologie de reconnaissance de visages, ainsi que leurs enjeux. La table 1.2 répertorie ces applications ainsi que leurs principales caractéristiques. Le sigle « BC » désigne la Base de Connaissance,

Compagnie	**Produit**	**Site internet**
Neven Vision, Inc. (anciennement Eyematic)	Mobile-i Face Recognition SDK	http://www.nevenvision.com
Identix, Inc.	FaceIt SDK	http://www.identix.com
HumanScan Co.	BioID SDK	http://www.bioid.com
Viisage	FaceFinder	http://www.viisage.com
Cognitec	FaceVACS	http://www.cognitec-systems.de
FaceKey Corp.	FaceKey	http://www.facekey.com
Visiphor (anciennement Imagis)	Visiphor BIE	http://www.imagistechnologies.com
Acsys Biometrics Corp.	Acsys FRS	http://www.acsysbiometricscorp.com
C-VIS GmbH	FaceSnap Recorder	http://www.c-vis.com
DreamMirh Co., Ltd.	MIRH Eye	http://www.dreammirh.com
VisionSphere Technologies, Inc.	UnMask Plus, It's Me, FaceCam	http://www.visionspheretech.com
Istituto Trentino di Cultura	SpotIt!	http://spotit.itc.it
ImageWare Systems, Inc.	IWS Products	http://www.iwsinc.com
Real User Corp.	Passfaces	http://www.realuser.com

TAB. 1.1 – *Principaux systèmes commerciaux de reconnaissance de visages (certains de ces sites internet peuvent avoir changé).*

qui contient toutes les images de personnes connues. La taille de la BC est qualifiée de *petite* si le nombre de personnes enregistrées n'excède pas une vingtaine, on parlera de taille *moyenne* pour un nombre d'individus inférieur à deux cents environ et de *grande* taille pour un nombre de personnes supérieur.

On peut classer les applications en deux grandes familles : celles *en monde fermé* (F), où tout visage-requête est enregistré dans la base de connaissance, et celles *en monde ouvert* (O), où des visages de personnes inconnues peuvent être présentées au système.

La reconnaissance peut consister en une tâche d'*identification* (Id), ou d'*authentification* (Au). Identifier un visage, c'est lui assigner une identité sans prendre en compte aucune information *a priori* sur sa classe d'appartenance présumée. En revanche, dans le contexte de l'*authentification* (*vérification*), toute personne se présentant au système revendique une certaine identité. Le processus consiste à vérifier qu'il s'agit bien de cette personne.

Le *niveau de sécurité* de l'application est directement lié aux conséquences d'une mauvaise

classification. Par exemple, on pourra considérer que si un système conçu dans un but de divertissement effectue une mauvaise classification, la portée de cette erreur sera moindre que si celle-ci avait amené à une usurpation d'identité pour une transaction financière.

Domaine	**Application**	**Tâche**		**Taille de BC**	**Niveau de sécurité**	**Monde**	
		Id	**Au**			**O**	**F**
Biométrie	Contrôle parental	X		petite	moyen		X
	Accès à son poste de travail		X	petite/ moyenne	haut	X	
	Sécurisation des transactions		X	petite	haut	X	
Indexation	Contenu personnel	X		petite	faible		X
	Contenu spécifique	X		moyenne	moyen	X	
	Contenu générique	X		grande	moyen	X	
Sécurité	Aide à la décision	X		moyenne/ grande	haut	X	
	Analyse de scènes *a posteriori*	X		grande	haut	X	
	Vidéosurveillance	X		grande	haut	X	
Divertissement	Interaction homme-machine	X		petite	faible		X

TAB. 1.2 – *Principales applications de la reconnaissance de visages. On détaille ici le domaine de mise en œuvre, la tâche (Identification/Authentification), la taille de la Base de Connaissance, le niveau de sécurité, ainsi que le contexte (monde Ouvert/Fermé).*

1.3.1 Biométrie

À ce jour, le moyen le plus répandu d'authentifier une personne de manière électronique repose sur l'utilisation d'un code personnel, composé de chiffres et/ou de lettres. Ce mode de vérification d'identité est relativement peu sûr, puisqu'il suffit de connaître le code de quelqu'un d'autre pour usurper son identité. De plus, la multiplication des transactions et des communications électroniques fait que l'usager moyen doit retenir un nombre croissant de codes : un pour retirer de l'argent depuis un distributeur automatique, un pour accéder à son ordinateur, plusieurs pour se connecter à des sites internet, etc. Si bien que le besoin de systèmes d'authentification conviviaux et sûrs se fait de plus en plus pressant. La *biométrie* consiste à authentifier une personne à l'aide de ce qu'elle est, et non pas de ce qu'elle sait ou détient. Parmi les méthodes de biométrie les plus efficaces, on peut citer celles basées sur l'étude de l'iris, de la rétine, ou des empreintes digitales. Mais ces techniques présentent l'inconvénient d'être *intrusives* (au sens

où elles nécessitent la coopération de l'utilisateur) ce qui réduit le champ des applications. De plus, leur déploiement repose sur l'utilisation d'un matériel dédié. À l'inverse, la technique de reconnaissance de visages est non-intrusive – on peut vérifier l'identité de quelqu'un sans même qu'il s'en rende compte – et un matériel de prise de vue courant (comme par exemple un appareil photographique numérique ou une webcam) suffit pour l'acquisition des données. Selon le rapport G-276 [GBM03] rédigé par la Business Communications Company, Inc., le marché global de la biométrie représentait 946 millions de dollars en 2002. Les prévisions font état d'un taux d'accroissement annuel moyen de 29,1% jusqu'en 2007, où le marché devrait représenter 3,4 milliards de dollars. C'est la reconnaissance d'empreintes digitales qui représente la part de revenus la plus importante, mais la reconnaissance faciale devrait avoir un taux d'accroissement annuel moyen plus important, avec 33,3% contre 29,2% pour les empreintes digitales. Les applications de biométrie listées en table 1.2 sont :

- le *contrôle parental* pour l'accès à la télévision passée une certaine heure par exemple. S'agissant de biométrie dans un cadre familial, on peut considérer que le monde est fermé et la base de connaissance de petite taille. Afin de rendre le système le plus convivial possible et notamment de ne pas avoir à solliciter la collaboration de l'utilisateur, on préférera dans ce contexte mettre en œuvre une tâche d'identification que d'authentification ;
- l'*accès à son poste de travail*, (bâtiment ou environnement personnel dans l'ordinateur de bureau). Le nombre de personnes autorisées d'accès étant limité, la taille de la BC est petite à moyenne. Il s'agit typiquement d'une application d'authentification en monde ouvert (puisque des imposteurs peuvent tenter d'accéder à la zone protégée). Il est possible d'amener l'utilisateur à coopérer, par exemple en lui demandant de placer sa tête à un emplacement précis, ce qui peut grandement simplifier la tâche de reconnaissance ;
- la *sécurisation des transactions*, dont les caractéristiques sont très semblables à celles de l'application ci-avant, sauf que le nombre de personnes autorisées est généralement plus faible.

1.3.2 Indexation de documents multimédia

Le volume sans cesse croissant d'information multimédia existante, tant sur internet que dans des organisations de conservation du patrimoine (telles que l'Institut National de l'Audiovisuel), rend nécessaire la mise à disposition de moyens performants d'*indexation* (classement) et de recherche parmi ces documents. L'identité des personnes représentées dans une image véhicule une information sémantique forte. Par exemple, si l'on reconnaît M. Jacques Chirac et M. Vladimir Poutine dans une image, il s'agit sans doute d'une photographie prise lors d'un évènement à caractère politique. Les applications d'indexation sont nécessairement effectuées *a posteriori* de la prise de vue et donc elles reposent généralement sur une tâche d'identification. Les principales applications d'indexation listées en table 1.2 sont :

- l'*indexation de contenu personnel*. Il peut s'agir par exemple de trier les photographies de vacances d'un utilisateur. On considère que le cercle des personnes potentiellement représentées (en général des amis) est fermé et limité (base petite à moyenne) ;
- l'*indexation de contenu spécifique*, tel que des Journaux Télévisés (JT). Dans cet exemple, il peut être utile de distinguer la séquence « plateau » de celles de « reportage ». Dans ce but, on peut utiliser l'outil de reconnaissance de visages, généralement conjointement avec d'autres techniques d'analyse d'images et de découpage en séquences. Ce système nous permet de vérifier les résultats par l'identification du présentateur. Le nombre de présentateurs potentiels de JT étant par définition limité, on considère que la BC est de

petite taille. Par contre, il peut y avoir des invités extérieurs et donc on est dans un contexte de monde ouvert ;

– l'*indexation de contenu générique*. On dispose d'une vidéo ou d'un ensemble d'images fixes, sans aucune information *a priori* sur son contenu. Le but est de déterminer si des personnes connues sont présentes dans ces documents. La BC est potentiellement grande et le monde ouvert.

1.3.3 Sécurité

Il est souvent souhaitable que la surveillance automatique de lieux publics ou privés se fasse de manière non intrusive, c'est-à-dire sans demander aux passants de décliner leur identité, ce qui évite d'installer des points de contrôle qui restreindraient les personnes dans leurs mouvements. On se place alors dans le cadre d'une tâche d'identification et non d'authentification. La reconnaissance de visages est l'une des techniques privilégiées dans ce contexte de surveillance, à cause de sa non intrusivité et de la relative facilité d'acquisition de photographies d'un visage qui rend la constitution de la base de connaissance assez aisée. Les principales applications de *sécurité* sont (*cf.* table 1.2) :

– l'*aide à la décision*. Un expert dispose de la photographie d'une personne, dont il cherche l'identité parmi un ensemble potentiellement grand de personnes connues. Il questionne le système qui doit lui fournir les noms des candidats possibles. C'est l'opérateur humain qui tranchera : soit aucun des candidats proposés ne correspond, et l'on pourra en déduire que le visage-requête n'est pas enregistré dans la base de connaissance, soit l'un de ces candidats convient et le visage-requête est identifié. Le monde est donc ouvert, et le nombre de personnes enregistrées va de quelques dizaines à plusieurs centaines ;

– la *vidéosurveillance*. Une caméra numérique est placée dans un lieu public ou privé, en intérieur ou en extérieur, et collecte des images en permanence. Dès qu'un visage est détecté, l'image correspondante est présentée au processus d'identification. La qualité de ces images peut être très mauvaise, et leur résolution très (trop) faible. Le temps de réponse du système doit obligatoirement être très rapide.

– l'*analyse a posteriori de contenu*. Les caractéristiques de cette application sont très semblables à celles de la vidéosurveillance ; la seule différence avec cette dernière est que l'on dispose généralement de plus de temps pour mener à bien l'identification.

1.3.4 Divertissement

Dans le domaine du *divertissement*, nous avons répertorié les interactions homme-machine, comme par exemple les jeux vidéos en réseau où chaque joueur, placé devant une webcam, est automatiquement reconnu. Son identité étant connue, on peut alors procéder à une configuration automatique de son environnement de jeu : il est par exemple possible de choisir automatiquement comme personnage du jeu l'avatar de cette personne parmi les avatars de l'ensemble des utilisateurs. Le nombre de joueurs potentiels étant limité, on peut considérer qu'il s'agit d'une application en monde fermé, où la taille de la base d'apprentissage est petite. Dans un but de convivialité, on privilégiera une tâche d'identification par rapport à l'authentification.

1.3.5 Applications visées dans le contexte de cette thèse

Dans cette thèse, nous nous intéresserons plus spécifiquement à la tâche d'identification, en monde fermé dans un premier temps (chapitre 4.), puis nous étendrons notre travail au monde

ouvert au chapitre 5. Nous cherchons à concevoir un système fiable pour des tailles de la base de connaissance petite à moyenne, c.-à-d. n'excédant pas une à deux centaines d'individus. Nous ne visons aucune des applications citées ci-dessus en particulier ; néanmoins, le développement d'un tel outil nous ouvre les domaines du divertissement, de la biométrie dans un cadre familial, de l'indexation de contenu personnel ou spécifique, et de l'aide à la décision dans le contexte de la sécurité. Les expérimentations reportées aux chapitres 4. et 5. seront conçues pour tester les performances des algorithmes proposés dans de tels *scenarii*.

1.4 La reconnaissance humaine de visages

L'aptitude à reconnaître ses semblables est l'une des capacités cognitives les plus importantes de la race humaine, indispensable à l'organisation en société. La reconnaissance d'une personne par le cerveau humain passe par l'utilisation de nombreuses informations visuelles (visage, posture, coupe de cheveux, etc.) ou non visuelles (voix, parfum, etc.). Le contexte est également un facteur important : on reconnaîtra plus rapidement un collègue sur son lieu de travail que dans la rue. Néanmoins, la perception d'un visage suffit généralement à sa reconnaissance : la plupart des humains sont capables, par l'étude d'une photographie de qualité suffisante, de reconnaître une personne connue, ou de distinguer deux visages différents (sauf s'il s'agit de vrais jumeaux), et ce en une fraction de seconde. Les premiers travaux portant sur la reconnaissance humaine de visages ont été menés dans le domaine de la neuropsychologie dès le début des années 1950. Depuis, notamment grâce aux avancées dans le domaine de l'imagerie médicale, des études de neurophysiologie sont venues compléter ces recherches. C'est avec un grand intérêt que les scientifiques cherchant à automatiser le processus suivent ces travaux, dans l'espoir de concevoir un algorithme capable de copier les prodigieuses facultés de reconnaissance du cerveau humain. Ainsi, Zhao *et al.* ont dressé dans [ZCPR03] un inventaire assez complet des principales études relatives à la reconnaissance humaine des visages.

1.4.1 La reconnaissance humaine de visages : un processus spécifique

Comme nous l'avons vu en introduction de cette thèse, la reconnaissance de visages constitue une tâche de classification plus complexe que la reconnaissance d'objets génériques [DC86, RML90, GT97]. Aussi serait-il plausible que cette tâche repose sur un mécanisme spécialisé et indépendant de ceux impliqués dans la reconnaissance d'autres objets. Plusieurs constatations viennent accréditer cette thèse [BK98]. Premièrement, le cerveau humain a la capacité de reconnaître un visage familier à partir de relativement peu d'information, comparé à d'autres objets. Le deuxième argument en faveur de cette hypothèse est l'existence d'une maladie (très rare) nommée *prosopagnosie* et telle que les sujets qui en sont affectés sont incapables de distinguer deux visages, ou de reconnaître visuellement des visages familiers, alors que leurs autres capacités de reconnaissance visuelle sont intactes. Les malades peuvent distinguer un visage d'un autre objet et détecter ses caractéristiques faciales, mais sont incapables de combiner ces éléments dans un but de reconnaissance. Chez certains d'entre eux, on a décelé une lésion cérébrale localisée dans une région spécifique du cortex visuel. Ce serait donc cette partie du cerveau qui serait impliquée dans la reconnaissance de visages, hypothèse que viennent accréditer des études neurophysiologiques montrant une activité accrue dans cette région lors du processus de reconnaissance de visages.

1.4.2 Utilisation des caractéristiques locales et globales

Les psychophysiciens et les neuroscientifiques ont cherché à déterminer si le cerveau humain se base plutôt sur l'étude de caractéristiques globales ou locales du visage pour sa reconnaissance. De nos jours, la plupart des chercheurs s'accordent à dire que les deux types d'informations sont utilisées [BHB98]. Si l'analyse de l'information des basses fréquences de l'image (reflétant sa globalité) sont suffisantes à la reconnaissance d'un visage très familier, les cheveux, la forme du visage, les yeux et la bouche sont reconnues comme étant des caractéristiques primordiales dans la perception et la mémorisation des visages. Beaucoup d'études montrent que l'importance du nez est insignifiante dans le cadre de la reconnaissance d'un visage de face, mais devient importante pour une pose de profil ou de trois-quarts, où l'on peut mieux appréhender sa forme. Dans le cadre de visages connus mais non familiers, l'œil humain prend en compte à la fois les caractéristiques intérieures (bouche, yeux) et extérieures (cheveux) à l'ovale du visage, tandis que pour la reconnaissance de visages familiers nous analysons essentiellement les caractéristiques intérieures. De plus, le fait que l'on soit doté de la capacité à reconnaître au premier coup d'œil le sujet d'une caricature tend à prouver le pouvoir discriminant des yeux, du nez et de la bouche. En effet, ce sont là les principaux traits croqués (et généralement exagérés) dans les caricatures [Per75]. Il a également été montré que la partie supérieure du visage a une plus grande importance que la partie inférieure.

1.4.3 Impact des différentes sources de variabilité

Il est également très intéressant d'étudier l'impact des différentes sources de variabilité sur les performances du système visuel humain en termes de reconnaissance des visages. Une personne ayant vu un visage en une seule occasion peut le reconnaître dans des conditions d'orientation, d'expression faciale ou de luminosité très différentes [TH96, BVB87, Mos93].

Des expérimentations ont montré qu'un visage familier éclairé par une source lumineuse situé sous le visage est plus difficile à identifier [JHC92]. Ainsi, la direction d'illumination influe sur notre aptitude à reconnaître un visage ; les conditions d'illumination les plus favorables consistent en un éclairage par le haut [HB96].

Selon des études neurophysiologiques, il semblerait que l'analyse de l'expression faciale se fasse par un processus indépendant de la reconnaissance du visage. Certains patients atteints de prosopagnosie peuvent catégoriser les expressions faciales alors même qu'ils sont incapables de reconnaître le visage les arborant. À l'inverse, certaines personnes souffrant d'un type particulier de syndrome organique du cerveau sont capables de reconnaître un visage, mais pas d'interpréter sa gestuelle faciale. De manière générale, l'expression faciale a peu d'influence sur nos capacités de reconnaissance, pour autant qu'elle reste raisonnable.

Par contre, la question de savoir si la reconnaissance est ou non indépendante de la pose n'est pas encore tranchée. Étant donné qu'une image de visage se présente généralement à l'endroit, le facteur le plus variable est la *rotation en profondeur*. On parle de rotation en profondeur pour des mouvements de type hochement de tête, ou négation. De nombreux scientifiques s'accordent à dire que la pose n'a d'influence significative sur la reconnaissance que si son amplitude est très importante. En effet, il est difficile de mettre en correspondance les deux profils d'un même visage, tandis que pour deux vues de trois quarts cette tâche est relativement aisée [HSA97]. Cette constatation tend à prouver que notre cerveau est capable d'utiliser, dans une certaine mesure, la symétrie du visage par rapport à son axe central vertical [TB98].

1.4.4 Discussion

Un certain nombre d'enseignements peuvent être tirés de ces études, dans le but de concevoir un système de reconnaissance automatique qui soit le plus efficace possible. Tout d'abord, les constatations données en section 1.4.1 ont amené un grand nombre de chercheurs à proposer des systèmes de reconnaissance de visages spécialement conçus pour cette tâche et non pas directement des applications de techniques plus générales de reconnaissance de formes. Deuxièmement, les éléments soulignés en section 1.4.2 plaident en faveur des techniques *hybrides* (au sens où elles sont basées sur l'étude conjointe des caractéristiques globales et locales des visages) et accordant une importance particulière aux caractéristiques issues de la partie supérieure du visage [ZCPR03].

Les études exposées en section 1.4.3 tendent à prouver l'importance de l'apprentissage pour la reconnaissance humaine : nous avons, depuis notre enfance, observé suffisamment de visages pour avoir complètement intégré leur symétrie, on est donc capable d'inférer une vue à partir de son opposée. À l'inverse, on a peu l'occasion de rencontrer dans la nature des sources d'éclairage provenant du sol ; par conséquent, il nous est difficile de reconnaître un visage dans ces conditions. Ces remarques montrent qu'il ne suffit pas de considérer les dissimilarités entre deux visages différents pour construire un outil de reconnaissance automatique performant : il faut également prendre en compte l'ensemble des variations possibles entre deux vues d'un même visage.

1.5 Difficultés inhérentes à la reconnaissance automatique

Les systèmes automatiques de reconnaissance de visages doivent rester invariants à tout facteur indépendant de l'identité du visage, même si ce facteur engendre des changements d'apparence du visage. Or, de nombreux facteurs, extérieurs au visage ou en lien avec sa nature intrinsèque, peuvent influer sur celle-ci. Les conditions de prise de vue, notamment l'angle sous lequel le visage est observé et la puissance des sources de luminosité, influent considérablement sur l'apparence d'un visage. Nous pouvons citer les propos de Moses *et al.* dans [MAU94] (traduit de l'anglais) : « Les variations entre des images d'un même visage dues à l'illumination et à l'angle de vision sont presque toujours plus importantes que les variations entre images dues à un changement de l'identité du visage ». L'expression faciale arborée par le sujet à l'instant où l'image est collectée, ainsi que d'éventuelles occultations partielles (une partie du visage est cachée par un autre objet, par exemple des lunettes) ainsi que le vieillissement peuvent également engendrer des changements d'aspect importants. Des facteurs individuels relatifs aux personnes à reconnaître, tels que leur sexe ou leur âge, ainsi que la taille de la base de connaissance, peuvent également avoir un impact sur les performances du système.

Dans cette section, nous passerons en revue ces principaux facteurs et étudierons leur impact. L'analyse des résultats expérimentaux obtenus dans le cadre des FRVT 2000 et 2002 [BBP01, PGM+03] et par Gross *et al.* dans [GSC01] nous permettra de tirer un certain nombre de conclusions. Le protocole FRVT, qui sera détaillé en section 1.7.3, a permis essentiellement de caractériser l'impact de la pose, des conditions d'illumination, du délai entre différentes prises de vue, de facteurs individuels et de la taille de la base. Dans [GSC01], Gross *et al.* fournissent une étude systématique de l'impact de différents paramètres sur les performances du système, variés de manière isolée ou conjointe. Les six facteurs considérés sont : la pose de la tête, les changements d'illumination, l'expression faciale, les occultations, l'intervalle de temps entre deux prises de vue et le sexe.

1.5.1 Les variations de pose

Un changement de l'angle d'inclinaison du visage engendre de nombreux changements d'apparence dans l'image collectée, (pour une position fixe du capteur). Nous nous intéresserons ici aux rotations du visage en profondeur (mouvement de type hochement de tête ou négation). En effet, on suppose que la phase préliminaire de normalisation du visage qui sera détaillée en section 1.6 permet de corriger d'éventuelles rotations dans le plan de l'image. Les rotations en profondeur engendrent deux types de difficultés. Tout d'abord, elles amènent des différences de profondeur qui, projetées sur le plan 2D de l'image, résultent en des déformations (étirement de certaines parties du visage et compactage d'autres régions). La forme du visage, et donc les distances entre caractéristiques faciales, varient. Secondement, elles peuvent mener à l'occultation de certaines parties du visage (par exemple, dans une vue de trois quarts, une partie du visage est cachée).

Si la pose du visage-requête diffère significativement de celle des visages enregistrés, les performances des systèmes de reconnaissance de visages sont affectées et les taux de reconnaissance baissent sensiblement, comme l'a mis en évidence le FRVT [BBP01]. En effet, selon le FRVT 2000, la rotation de la tête n'entraîne pas de baisse des taux de reconnaissance significative jusqu'à $\pm 25°$, alors qu'à partir de $\pm 40°$ on constate une chute des performances.

Dans [GSC01], Gross *et al.* ont montré que, si le seul facteur de variation entre l'image enregistrée et l'image-requête est une rotation en profondeur de la tête inférieure à 30°, les taux de reconnaissance des systèmes (statistiques) actuels sont de l'ordre de 90%. Des rotations plus importantes engendrent une forte baisse des performances. Pour beaucoup d'applications telles que la biométrie, les angles de rotation sont généralement inférieurs à 30° et donc les performances des algorithmes actuels sont très intéressantes. Par contre, pour d'autres applications telles que la vidéosurveillance, il n'est pas suffisant de garantir d'excellentes performances avec un angle de rotation inférieur à 30°. En effet, les caméras de surveillance sont souvent localisées en hauteur et, si elles sont en intérieur, proches des coins des pièces, ce qui implique naturellement des angles de prise de vue en dehors de ces limites. Gross *et al.* ont également mis en évidence le fait qu'un modèle construit à partir de poses frontales présente une meilleure capacité de généralisation à d'autres poses qu'un modèle construit à partir de poses non frontales.

Il est parfois nécessaire, pour certains types de techniques et dans le contexte de certaines applications, de construire plusieurs modèles de reconnaissance (un par type de pose). En phase de reconnaissance, il faudra alors utiliser un *classifieur de pose* [3] en amont de la reconnaissance, de manière à ce que tout visage-requête ne soit comparé qu'au modèle de sa pose.

1.5.2 Les changements d'illumination

L'intensité et la direction d'illumination lors de la prise de vue influent toutes deux énormément sur l'apparence du visage dans l'image. Dans la plupart des applications réelles, des changements dans les conditions d'illumination sont néanmoins inévitables, notamment lorsque les vues sont collectées à des dates différentes, en intérieur ou en extérieur. Étant donné qu'un visage humain est un objet intrinsèquement 3D, des changements d'illumination peuvent faire apparaître sur le visage des ombres accentuant ou, au contraire, masquant certaines caractéristiques faciales (*cf.* figure 1.2).

L'évaluation du FRVT [BBP01] conclut que des changements importants dans les conditions d'illumination peuvent mener à des baisses considérables dans les taux de reconnaissance. En

3. Module permettant de caractériser l'angle de prise de vue du visage, à partir par exemple des positions des caractéristiques faciales.

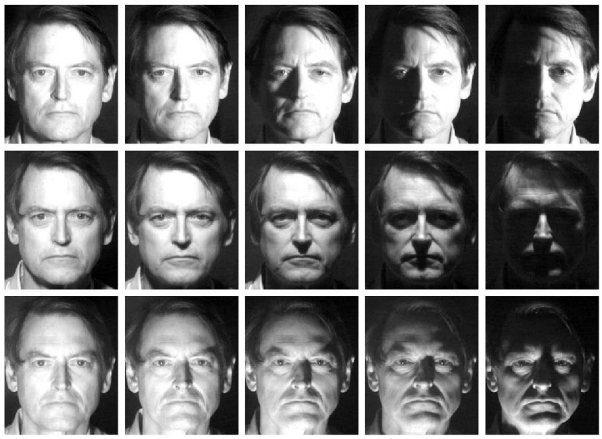


FIG. 1.2 – *Extrait de [MAU94]. Effets de variations dans les directions d'illumination sur l'apparence d'un visage.*

effet, si la plupart des systèmes de reconnaissance de visages sont stables à des changements raisonnables des conditions d'illumination en intérieur, en extérieur on note des déficits de reconnaissance de l'ordre de 40% avec les meilleurs outils testés par le FRVT. Ces baisses de taux de reconnaissance peuvent être dues à des variations de la somme de luminosité réfléchie par la peau et/ou à des réglages effectués automatiquement par la caméra pour garantir une bonne qualité d'image (par exemple la correction gamma, le contraste et les propriétés d'exposition).

Gross *et al.* [GSC01] ont étudié de manière isolée l'impact du facteur de réflexion de la peau, grâce à une étude menée sur la base PIE, décrite en annexe A. Les résultats obtenus montrent que les algorithmes de reconnaissance de visages actuels sont en fait robustes aux effets de changements d'illumination purs (déjà en partie corrigés par l'égalisation d'histogramme appliquée lors de la phase de normalisation). Ce seraient donc les ajustements non linéaires des caméras qui engendreraient en grande partie les baisses de performance des techniques linéaires observées dans le cadre du FRVT, et imputées à tort dans un premier temps à des changements d'illumination. Par conséquent, lorsque l'on peut contrôler les conditions de prise de vue (biométrie, sécurité), l'on ne saurait trop recommander de contrôler les réglages de la caméra. Par contre, puisqu'il est difficile de corriger les effets des réglages automatiques, il nous faudra faire avec lorsque l'on ne peut les contrôler (applications d'indexation par exemple). Gross *et al.* remarquent également que les performances des algorithmes baissent sensiblement lorsque les conditions d'illumination extrêmes engendrent une occultation d'une partie du visage. Par conséquent, le couplage de variations d'illumination et de variations de pose constitue une difficulté importante.

1.5.3 Les expressions faciales

Les visages sont des objets non rigides. L'expression faciale de l'émotion, surtout combinée avec la parole, peut produire des changements d'apparence importants des visages. Le nombre de configurations possibles se compte en milliers. L'influence de l'expression faciale sur la reconnaissance est donc difficile à évaluer. Puisque l'expression faciale affecte la forme géométrique et les positions des caractéristiques faciales, il semble logique que les techniques globales ou hybrides y soient plus robustes que la plupart des techniques géométriques. Gross *et al.* [GSC01] ont étudié l'impact de changements d'expressions faciales sur la reconnaissance, grâce à la base des expressions faciales de Kohn-Kanade [KCT00]. Leurs résultats expérimentaux montrent que

les algorithmes sont relativement robustes aux changements d'expression faciale, à l'exception des cas extrêmes engendrant d'importantes déformations de la bouche (telles que le cri) et le rétrécissement ou la fermeture complète des yeux.

Il peut donc être utile de repérer en amont de la reconnaissance ces expressions problématiques. Si l'on est capable de catégoriser l'expression faciale du visage-requête, deux approches sont possibles. Soit plusieurs modèles de visages ont été appris (un par catégorie d'expression faciale), et l'on peut alors comparer directement le visage de test à la base des visages arborant la même expression. Soit on utilise une technique *générative* qui, grâce à l'utilisation d'un modèle de visages suffisamment précis, nous permet de transformer le visage-requête de manière à ce qu'il se présente dans des conditions moins difficiles. Il existe de nombreux systèmes de reconnaissance des expressions faciales. Pour des études bibliographiques relativement complètes et récentes, se reporter à [PR00, FL03a]. Certaines approches [BLFM03] consistent à classer les émotions en sept expressions basiques : neutre, colère, dégoût, peur, joie, tristesse, surprise. D'autres, constatant que ce genre d'émotions basiques n'est que rarement observé dans le cadre d'applications réelles, ont mis en place des systèmes capables de reconnaître des changements plus subtils d'expression [TKC02].

1.5.4 Les occultations partielles

Le visage peut être partiellement masqué par des objets dans la scène, ou par le port d'accessoires tels que des lunettes de soleil. Les occultations peuvent donc être intentionnelles ou non. Dans certains cas, il peut s'agir d'une volonté délibérée de contrecarrer la reconnaissance (dans le contexte de la vidéosurveillance par exemple). Excepté dans le contexte de la biométrie ou du divertissement, les systèmes proposés doivent être non intrusifs, c.-à-d. qu'on ne peut pas compter sur une coopération du sujet. Par conséquent, il est important de savoir reconnaître des visages partiellement occultés. Gross *et al.* ont évalué dans [GSC01] l'impact du port de lunettes de soleil, et d'un cache-nez occultant la partie inférieure du visage, par le biais de l'utilisation de la base AR (*cf.* annexe A). Leurs résultats expérimentaux montrent que les performances des algorithmes testés sont en général faibles dans ces conditions. De plus, les différents algorithmes présentent un comportement différent vis-à-vis des occultations. Nous reviendrons sur ce point en section 2.4.

1.5.5 Le vieillissement et les changements d'aspect

Les visages changent d'apparence au fil du temps. Les modifications concernent la tension des muscles, l'apparence de la peau (apparition de rides), le port de lunettes, éventuellement le maquillage ou la présence d'une frange occultant une partie du front.

Gross *et al.* [GSC01] utilisent la base AR (*cf.* annexe A) pour déterminer l'impact de ces facteurs. Sur la base AR, où le délai entre deux prises de vue est seulement de deux semaines, la baisse des taux de reconnaissance est estimée à 20%.

Dans le FRVT 2000, les effets du temps ont été mesurés à l'aide des vues *duplicate* de la base FERET : celles-ci sont comparées aux vues FA, ce qui permet d'établir un taux de reconnaissance (*cf.* annexe A). Les taux fournis par les meilleurs algorithmes sont de 63% sur les vues duplicate I, et 64% pour les vues duplicate II, contre 58% et 52% pour l'évaluation de 1996. Récemment, les systèmes de reconnaissance ont donc réalisé d'énormes progrès pour gérer au mieux le délai de temps entre deux prises de vue. Néanmoins, bien que l'intervalle de temps entre les vues FA et les vues duplicate I ne soit pas nécessairement important, les systèmes ont du mal à reconnaître

ces dernières. Cela provient certainement des changements dans les conditions de prise de vue, et non d'un vieillissement des visages.

Dans le cadre de l'évaluation FRVT 2002 [PGM+03], la baisse des taux de reconnaissance (des meilleurs algorithmes testés) a été estimée à 5% par année d'écart entre l'image de référence et l'image à reconnaître.

À notre connaissance, les effets du développement de la personne (par exemple le passage de l'enfance à l'adolescence) restent à ce jour inexplorés.

1.5.6 Les facteurs individuels

Le sexe de la personne à reconnaître, son âge ainsi que son groupe ethnique, peuvent également influer sur les performances de l'algorithme de reconnaissance.

Étudions tout d'abord les différences d'apparence du visage en fonction du sexe [BY98]. Outre des différences de forme du visage, les sourcils des hommes sont généralement plus épais, et la région basse de leur visage est généralement plus texturée à cause de la barbe. Dans les visages de femme, la distance entre les yeux et les sourcils est généralement plus importante, le nez plus petit, et le menton plus étroit. Intuitivement, les visages de femme devraient être plus difficiles à reconnaître, notamment à cause du maquillage. Les résultats expérimentaux reportés par Gross *et al.* viennent contredire cette hypothèse, puisque les algorithmes évalués reconnaissent plus facilement les femmes que les hommes (dans les bases AR et FERET), et ceci en présence de différentes sources de variation et malgré une représentation équivalente des deux sexes dans la base d'apprentissage. *A contrario*, les expérimentations reportées dans le cadre du FRVT 2002, plus significatives statistiquement car portant sur plus de sujets, ont montré que les taux d'identification pour les hommes étaient de 6% à 9% meilleurs que ceux obtenus pour les femmes. Cette contradiction entre les deux évaluations met une fois de plus en évidence le fait que les performances des systèmes sont très liées aux bases utilisées pour leur évaluation.

Lors du FRVT 2002, l'impact de l'âge des sujets sur les performances du système a également été étudié. Selon les conclusions reportées, plus la personne est âgée, plus les taux d'identification sont importants, avec une hausse d'environ 5% tous les dix ans (la pyramide des âges de la base utilisée s'étale de 18 à 77 ans). Cette constatation pourrait être expliquée par le fait que les visages des plus anciens sont plus texturés, du fait de la présence de rides, et présentent donc davantage de signes distinctifs que les plus jeunes.

1.5.7 L'impact de la taille de la base

C'est lors du FRVT 2002 que l'impact de la taille de la base sur les performances du système a été étudié pour la première fois, grâce à la très grande taille de la base HCInt utilisée [PGM+03]. Les meilleurs systèmes fournissent des taux de reconnaissance de l'ordre de 85% pour 800 personnes, 83% pour 1600 individus, et 73% pour 37437 personnes. Selon les conclusions de ce rapport, les performances décroîtraient de manière log-linéaire en fonction de la taille de la base.

1.5.8 Conclusion

Dans cette partie, nous avons listé les principales difficultés rencontrées dans le contexte de la reconnaissance automatique de visages. Nous avons montré que, parmi les facteurs influant le plus sur les performances du système, on compte les changements de pose, les occultations partielles des visages, et l'intervalle de temps entre deux prises de vue. Les baisses des performances sont d'autant plus sensibles que ces facteurs sont présents simultanément. D'autres facteurs, qui ont pourtant également une grande influence sur les taux de reconnaissance, ont fait l'objet de moins

d'études. On peut relever parmi ceux-ci la rotation et les variations d'échelle des visages dans les images. Le processus de normalisation des visages, détaillé en section suivante, vise à réduire l'amplitude de ces variations.

1.6 Détection et segmentation du visage dans l'image

Dans la grande majorité des algorithmes de reconnaissance de visages, le modèle est construit à partir d'images de visages contenant uniquement la région faciale. La prise en compte du fond de l'image reviendrait à introduire du bruit dans le modèle. Chen *et al.* [CLLH01] ont montré en 2001 que les techniques statistiques notamment seraient, dans le cas contraire, plus influencée par le fond que par la région faciale en elle-même. Généralement, on évite de prendre en compte des cheveux, car un changement de coiffure pourrait faire chuter drastiquement les taux de reconnaissance. Ces considérations sont en accord avec les études biologiques décrites en section 1.4 qui montrent que, pour reconnaître des visages familiers, l'œil humain s'attache plus aux caractéristiques faciales situées à l'intérieur de l'ovale du visage qu'à l'extérieur de celui-ci. Le visage doit donc être précisément segmenté et extrait de l'image. Pour cela, un processus de prétraitement, illustré en figure 1.3, doit être appliqué. Ce processus se décompose en trois étapes. La première consiste à détecter le visage dans l'image. Dans une seconde phase, on met en œuvre à l'intérieur de la région ainsi délimitée un module de détection des caractéristiques faciales, c.-à-d. des yeux, du nez, et de la bouche. Puis, on mène une étape de *normalisation* : à l'aide des positions des caractéristiques faciales, tous les visages sont centrés et alignés de la même manière dans les images correspondantes.

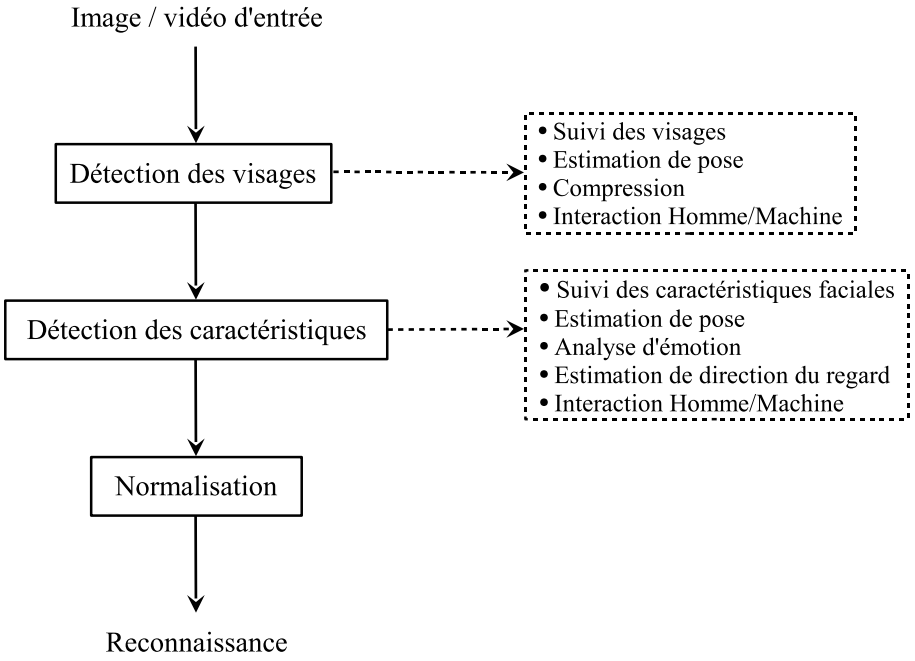


FIG. 1.3 – *Prétraitements des images en amont de la reconnaissance. Les différentes étapes du prétraitement peuvent également servir à d'autres applications (encadrés en pointillés).*

Notons que certaines techniques reposent sur la mise en œuvre simultanée de la détection des visages et de la détection des caractéristiques faciales. Dans la suite de cette section, nous

fournissons plus de détails concernant les différentes phases du prétraitement.

1.6.1 Détection des visages

Étant donnée une image, le module de détection a pour objectifs de décider si cette image contient un (ou plusieurs) visages, et de donner sa (leur) localisation(s) dans l'image par le biais d'une boîte englobante. Nous adoptons la définition de la détection proposée par Garcia et Delakis dans [GD04] : un visage est considéré comme correctement détecté si la taille de la fenêtre n'excède pas de plus de 20% la taille réelle de la région faciale, et qu'elle contient les yeux, le nez, et la bouche.

Deux principales statistiques caractérisent les performances d'un système de détection : le *taux de détection*, c.-à-d. le pourcentage de visages correctement détectés, et le taux de *faux positifs*, correspondant à des détections dans des régions ne contenant pas de visages. Un processus performant est associé à la fois à un taux de détection important et à un taux de faux positifs faible. Contrairement à la reconnaissance de visages, il existe des bases de référence utilisées pour l'évaluation de la plupart des algorithmes de détection de visages. Un état de l'art des principales techniques de détection de visages, ainsi qu'une comparaison des performances de celles-ci, est donnée en annexe B. Notons que le détecteur de visages reportant les meilleures performances sur les bases d'évaluation usuelles est a été introduit en 2004 par Garcia et Delakis dans [GD04]. Celui-ci fournit sur la base CMU un excellent taux de détection de 90,3%, avec un nombre de faux positifs (8) très faible, soit le meilleur ratio des techniques de l'état de l'art (voir table B.1) de l'annexe B. Il est capable de détecter des visages tournés de $\pm 20°$ dans le plan, et jusqu'à $\pm 60°$ en profondeur, ce qui est largement suffisant dans la plupart des applications réelles (*cf.* section 1.3).

1.6.2 Détection des caractéristiques faciales

Les traits perceptuellement les plus importants dans un visage sont les yeux, le nez, et la bouche. La détection de ces caractéristiques est une étape-clé du prétraitement des visages pour leur reconnaissance. En effet, les approches locales et hybrides ont intrinsèquement besoin de leurs coordonnées, et les techniques statistiques usuelles (telles que les *eigenfaces* et les *fisherfaces*, détaillées au chapitre suivant), reposent sur un alignement correct de tous les visages dans les images. La détection des caractéristiques faciales constitue un préalable indispensable à cette phase de normalisation, détaillée en section suivante. La recherche est restreinte à l'intérieur de la boîte englobante renvoyée par le module de détection du visage.

Outre la normalisation des visages pour la reconnaissance, la détection de caractéristiques faciales sert également dans le cadre de nombreuses autres applications telles que leur suivi, l'estimation de pose et de direction du regard, l'analyse d'émotion, et les interactions homme/machine (*cf.* figure 1.3). Les enjeux étant importants, de nombreuses approches ont été proposées durant la dernière décennie. Un état de l'art de ces méthodes est proposé en section B.2 de l'annexe B (p. 180). Notons qu'il existe à ce jour des techniques permettant de détecter rapidement et précisément les caractéristiques faciales jusqu'à $\pm 60°$ dans le plan de l'image, et $\pm 30°$ en profondeur [DG05].

1.6.3 Normalisation

La phase de normalisation permet, grâce aux positions des caractéristiques faciales précédemment détectées, d'aligner tous les visages de la même manière, et ainsi de créer des images de visages normalisées. Généralement, la normalisation consiste en une rotation du visage dans

l'image de manière à ce que l'axe interoculaire soit horizontal, suivie d'un centrage du visage dans l'image et d'un découpage de l'image de manière à ne retenir que la région faciale. Tous les visages doivent être représentés à la même échelle, et les images normalisées sont toutes de même taille. Les images de couleur sont ramenées en niveau de gris. La plupart du temps, un algorithme d'*égalisation d'histogramme* est appliqué. Celui-ci consiste à harmoniser la répartition des niveaux de luminosité de l'image, de manière à tendre vers un même nombre de pixels pour chacun des niveaux de gris de l'histogramme. Cette opération vise à harmoniser les valeurs de pixels des différentes images, afin de gommer les dissimilarités dues à des différences de conditions d'illumination par exemple. Le processus de normalisation utilisé dans le cadre de cette thèse est précisément détaillé en annexe C.

1.7 L'évaluation des performances des systèmes de reconnaissance automatique

Étant donnés le nombre et la variété des algorithmes de reconnaissance de visages proposés ces dernières années (voir figure 1.1), il est très important de définir des protocoles expérimentaux permettant leur évaluation. Mais la multiplicité des applications visées (voir table 1.2) rend cette évaluation difficile. En effet, il est nécessaire de disposer de données qui soient très semblables à celles rencontrées en situation réelle, et le type de données varie énormément d'un type d'application à un autre. Par exemple, dans le cadre de la biométrie, les conditions de prise de vue sont la plupart du temps contrôlées, alors que ce n'est pas le cas en général pour l'indexation ou la vidéosurveillance. De plus, la mesure des performances doit se faire selon une méthodologie prenant en compte le coût des erreurs, en fonction du niveau de sécurité des applications (voir table 1.2), comme nous le détaillerons en section 1.7.1. Par conséquent, tout mode d'évaluation est dépendant de l'application visée. Il existe de nombreuses bases pour l'évaluation, chacune ayant des attributs différents. Les performances des algorithmes sont très dépendantes des attributs des bases (nombre de personnes, variabilité des vues, etc.), et il n'existe pas de base unique d'évaluation. Les bases les plus utilisées sont décrites en annexe A (p. 169).

Néanmoins, durant la dernière décennie, des efforts ont été déployés pour définir des protocoles d'évaluation standardisés. La série des évaluations FERET [RPM98, PWHR98, PMRR00] menées par le National Institute of Standards and Technology (NIST) ont permis une comparaison de neuf systèmes, proposés par des institutions et des entreprises, dans des contextes d'authentification et d'identification. À ces évaluations ont succédé celles du Face Recognition Vendor Test (FRVT) [BBP01, PGM+03]. En parallèle, le protocole d'évaluation XM2VTS a été introduit essentiellement pour l'évaluation d'algorithmes de reconnaissance multimodaux (voix + visages) et 3D, et ceci pour l'authentification uniquement. Nous ne détaillerons pas ce protocole dans cette section. Pour plus de détails se référer à [MMK+99].

1.7.1 Les statistiques de mesure de la performance

L'estimation des performances de systèmes de reconnaissance de visages est une tâche difficile. Un certain nombre de conseils pour définir un bon système d'évaluation sont prodigués dans [MW02]. Dans le cadre de cette thèse, nous suivrons les lignes directrices définies dans ce rapport. Notamment, nous utiliserons systématiquement pour l'évaluation des bases test disjointes des bases de connaissance et d'apprentissage. Les performances des systèmes sont mesurées par des statistiques différentes selon que l'application évaluée est une tâche d'identification ou d'authentification. Les statistiques utilisées dans un contexte d'identification varient selon qu'il

s'agit d'une identification en monde fermé ou ouvert.

1.7.1.1 Statistiques utilisées pour l'identification en monde fermé

Pour l'*identification en monde fermé*, la première mesure à calculer est le pourcentage de visages de la base de test reconnus (c.-à-d. correctement assignés à leur classe d'appartenance). Cette statistique est couramment appelée *taux de reconnaissance*. Dans le cadre de l'identification toujours, mais dans le contexte d'applications semi-automatiques d'aide à la décision (voir figure 1.2), on peut également tracer les courbes *Cumulative Match Characteristic* (CMC). Ces courbes permettent d'obtenir les taux de reconnaissance cumulés en fonction du rang de reconnaissance. Un visage est reconnu au rang r si une vue du même visage est parmi les r plus proches voisins, au sens de la distance Euclidienne. Un visage correctement classé au rang $r = 1$ est automatiquement reconnu. Par contre, le fait qu'un visage soit reconnu au rang $r > 1$ n'assure pas sa reconnaissance automatique ; néanmoins, dans le cadre d'applications d'aide à la décision, où l'on considère qu'un opérateur humain saura sélectionner parmi les r plus proches voisins l'identité correcte, les courbes CMC constituent un indicateur de performance adapté.

1.7.1.2 Statistiques utilisées pour l'identification en monde ouvert

Pour évaluer les performances d'un algorithme dans le contexte d'une identification en monde ouvert, nous utiliserons deux bases de test. La première est composée de visages enregistrés dans la base de connaissance. La seconde base ne contient que des visages de personnes non enregistrées. Le processus de reconnaissance peut alors être décomposé en deux étapes. Dans une première phase, on effectue un *filtrage* des visages selon qu'ils appartiennent ou non à la base de connaissance. Pour cela, il nous faut définir une mesure de similarité entre l'image-requête et la base de connaissance. Si cette mesure de similarité est supérieure à un certain seuil fixé θ, on considère que le visage-requête est effectivement enregistré dans la base de connaissance. Sinon, on le rejette comme ne faisant pas partie de la base. Pour certaines applications de vidéosurveillance par exemple, il suffira qu'un visage passe cette phase de préfiltrage pour qu'une alarme soit déclenchée. Deux types d'erreur sont possibles. Un visage enregistré dans la base de connaissance peut être rejeté comme ne lui appartenant pas : il s'agit d'une erreur d'identification de *type I* appelée *faux rejet*. *A contrario*, un visage non enregistré dans la base peut être injustement déclaré comme lui appartenant : on parlera alors d'erreur de *type II* ou de *fausse alarme*. Le taux de faux rejet est évalué sur la première base, tandis que le taux de fausses alarmes est évalué sur la seconde base. Un seuil θ trop petit entraînera l'apparition d'un grand nombre de faux rejets, tandis qu'un θ trop grand engendre un taux de fausses alarmes important. Le seuil θ doit par conséquent être choisi de manière à garantir un bon compromis en erreur de type I et erreur de type II, pour le niveau de sécurité requis par l'application. Dans une seconde étape, on ne conserve que les visages ayant passé le préfiltrage et l'on peut donc se considérer en monde fermé. Les mêmes statistiques que précédemment peuvent être utilisées.

1.7.1.3 Statistiques utilisées pour l'authentification

Dans un contexte d'authentification (vérification), on appelle *client* une personne effectivement enregistrée dans la base, et qui donne son vrai nom au système (et devrait donc être accepté). Par opposition, on appelle *imposteur* un individu se réclamant d'une autre identité que la sienne (qui cherche à usurper l'identité de quelqu'un d'autre), et devrait donc être rejeté. Deux types d'erreurs peuvent survenir : le rejet d'un client (erreur de *type I*), ou l'acceptation

d'un imposteur (erreur de *type II*). Le pourcentage d'erreurs de type I sur la base de test est appelé *taux de faux rejet* (TFR), le ratio d'erreurs de type II est le *taux de fausses acceptations* (TFA). En général, la décision d'accepter ou de refuser un individu est prise sur la foi d'un score de mise en correspondance, en fonction d'un paramètre de seuil θ. Si le score est supérieur à θ, alors on autorise l'accès à l'utilisateur. Par contre, si le score est inférieur à θ, on lui refuse l'accès. Plus le seuil θ est grand, plus le taux de fausses acceptations est important. Un seuil trop faible, au contraire, engendre de nombreux faux rejets. Le paramètre θ doit être ajusté de manière à garantir un bon compromis entre TFR et TFA, pour le niveau de sécurité désiré. La statistique la plus simple pour mesurer la performance d'un algorithme dans le contexte de la vérification est le *taux d'erreur égale*. Pour calculer celui-ci, on règle le paramètre θ de manière à ce que TFR=TFA. Le taux d'erreur égale est alors la valeur de TFR (ou TFA) en ce point. L'évaluation passe également par le tracé de statistiques plus complexes, telles que les courbes Receiver Operating Characteristic (ROC). Cette courbe donne le TFR en fonction du TFA. Elle est tracée de manière paramétrique en fonction des valeurs de θ.

1.7.2 Le protocole FERET

La première évaluation FERET a été menée en août 1994. Ce protocole a été le premier conçu pour mesurer la performance d'algorithmes de reconnaissance automatique des visages. Depuis, deux autres évaluations ont eu lieu, l'une en mars 1995, l'autre entre septembre 1996 et mars 1997. Les détails des expérimentations menées et les résultats sont reportés dans [RPM98, PWHR98, PMRR00]. Les algorithmes ont été testés à la fois dans un contexte d'identification et d'authentification. La base de visages utilisée est la base FERET, décrite en annexe A. Les statistiques d'évaluation utilisées sont détaillées dans la section précédente.

Les neufs algorithmes proposés ont des performances variables en fonction des images-requêtes testées, et de l'application. Selon que la tâche est l'identification ou l'authentification, le meilleur algorithme n'est pas forcément le même. Les évaluations de FERET ont également permis de mettre en évidence les principaux facteurs pouvant influer sur les performances des algorithmes, tels que les changements de pose ou d'illumination, qui seront détaillés en section 1.5. Elles ont également mis en lumière les performances des différents algorithmes évalués ; plus de détails seront donnés en section 2.4, après que ces techniques aient été détaillées.

1.7.3 Les évaluations FRVT

Les protocoles FERET visaient à évaluer des prototypes mis à disposition par les laboratoires. Le développement à partir de 1997 de logiciels commerciaux (voir table 1.1) a rendu nécessaire l'évaluation de ces logiciels. C'est le but des FRVT 2000 et 2002.

En 2000, cinq compagnies ont participé à cette évaluation. Un protocole expérimental similaire à celui de FERET (session de Septembre 1996) a été utilisé, à la différence près que la base utilisée en 2000 comptait beaucoup plus d'images.

En 2002, dix participants (industriels) ont été évalués, entre juillet et août 2002. Le but du FRVT 2002 était de fournir des mesures de performances sur des images proches de celles rencontrées dans les applications réelles. La base de visages utilisée était la base High Computational Intensity (HCInt). La base HCInt contient 121589 images de 37437 personnes différentes. Les images sont issues du Département d'État Américain aux archives des visas pour les ressortissants Mexicains non immigrants. Cette base n'est évidemment disponible que pour les organisations ayant participé au FRVT. Les performances des algorithmes ont été mesurées dans le contexte de l'identification et de la vérification.

1.7.4 Discussion

Les évaluations menées dans le cadre de FERET et du FRVT ont un triple intérêt. Premièrement, elles ont permis dans une certaine mesure l'uniformisation des statistiques de mesure des performances utilisées. Deuxièmement, chacune d'entre elles fournit une évaluation des systèmes qui lui sont contemporains, et cela nous permet de suivre les évolutions des systèmes existants, et les progrès réalisés dans le domaine de la reconnaissance de visages. Nous y ferons régulièrement allusion dans la suite de cette thèse et notamment au chapitre suivant d'état de l'art. Enfin, elles ont permis de mettre en évidence les principales sources de difficulté pour les algorithmes de reconnaissance de visages, et ainsi de dessiner les futures voies de recherche.

Néanmoins, la base FERET ne constitue pas la base de référence unique pour l'évaluation des algorithmes de reconnaissance de visages, d'abord parce qu'elle ne contient pas suffisamment d'images par personne pour garantir un apprentissage efficace pour certains systèmes. De plus, elle ne permet pas une étude de sensibilité à certains facteurs, alors que d'autres bases, elle, sont conçues pour cela. Par exemple, les expressions faciales sont étiquetées avec beaucoup plus de précision dans les bases Yale ou PF01 (*cf.* annexe A), que nous préférerons pour évaluer la robustesse à l'expression. La base HCInt, quant à elle, n'est pas distribuée. Les évaluations FERET et FRVT n'ont donc pas permis d'imposer une base unique pour l'évaluation des algorithmes.

1.8 Conclusion

Ce chapitre nous a permis de définir plus précisément la problématique de l'identification des visages étudiée dans cette thèse, ainsi que le champ des applications possibles et de leurs enjeux. Nous avons tiré un certain nombre d'enseignements du processus de reconnaissance humaine des visages. Après avoir mis en lumière les principales difficultés inhérentes à la reconnaissance automatique de visages, nous avons présenté le processus de normalisation des visages dans les images, conçu et déployé de manière à essayer de réduire certaines de ces difficultés. Enfin, nous avons détaillé les principaux protocoles d'évaluation des (nombreux) systèmes de reconnaissance automatique proposés dans la littérature. La présentation des principales techniques utilisées dans ces systèmes fait l'objet du chapitre suivant.

Chapitre 2

État de l'art des techniques de reconnaissance automatique de visages

2.1 Introduction

Nous avons vu au chapitre 1. que la reconnaissance automatique de visages est un domaine de recherche très actif (voir figure 1.1). Nous introduisons ici les principales approches basées sur l'analyse des images de visages en niveaux de gris. Du fait du nombre et de la diversité des techniques proposées, la liste des méthodes détaillées dans ce chapitre n'est pas exhaustive.

Comme nous l'avons évoqué en introduction de cette thèse, la tâche de reconnaissance de visages peut se décomposer en deux étapes : l'extraction de caractéristiques et leur classification (voir figure 1). La première étape vise à fournir une *représentation* des visages sous la forme de *signatures*, tandis que la seconde étape constitue une phase de *mise en correspondance* de ces signatures. Le choix de la technique de classification est en général très dépendant du type des caractéristiques extraites ; il sera détaillé au fil de ce chapitre pour chacune des techniques présentées.

Ce chapitre est organisé comme suit. En section 2.2, nous étudierons les méthodes dites *globales*, au sens où les caractéristiques sont directement extraites depuis la totalité des pixels (en niveaux de gris) de l'image. Les approches *locales*, c'est-à-dire basées sur l'étude de caractéristiques extraites localement de différentes régions du visage, ainsi que les approches *hybrides* (alliant représentations globale et locale) font l'objet de la section 2.3. Enfin, la section 2.4 fournit une comparaison des performances de la plupart des méthodes présentées.

2.2 Les approches globales

Dans cette section, nous allons passer en revue les principales approches globales de reconnaissance de visages. Dans un premier temps, nous présenterons la technique d'étude des corrélations. Puis, nous détaillerons les principales approches de *projection statistique*, basées pour la plupart sur les techniques d'Analyse en Composantes Principales, d'Analyse Discriminante Linéaire, ainsi que d'Analyse en Composantes Indépendantes. Nous présenterons ensuite les méthodes reposant sur les Modèles Actifs d'Apparence, les réseaux de neurones et les Machines à Vecteurs de Support.

2.2.1 La corrélation

Les techniques globales les plus directes reposent sur l'utilisation d'un critère de similarité calculé entre les valeurs de pixels des images à comparer. L'une de ces méthodes consiste à mettre en correspondance le visage-requête avec l'exemple auquel elle est le plus corrélée [BP94]. Le critère retenu est généralement la corrélation croisée normalisée [Bar81, BP93]. La normalisation porte sur les distributions des pixels des deux images, dont les moyennes et les variances sont ramenées à des mêmes valeurs. Si cette normalisation dote la technique d'une certaine robustesse aux différentes sources de variation, la corrélation (tout comme la distance Euclidienne) est moins performante que la distance L_1 [BM95] et reste très sensible aux changements d'illumination, d'échelle et de pose du visage. Les résultats les plus robustes aux variations d'illumination utilisent la corrélation entre les valeurs de gradient (somme des gradients horizontaux et verticaux) des pixels des images. Afin d'accroître la robustesse aux changements d'échelle et de diminuer le temps de calcul, Burt a proposé en 1988 une approche hiérarchique [Bur88]. Beymer [Bey94] a apporté une amélioration susceptible de rendre la méthode plus robuste aux changements de pose. Il s'agit d'inclure une phase préliminaire de caractérisation de la pose. L'image-requête est alors comparée uniquement aux exemples correspondant à la même pose. Notons que cette technique est très peu tolérante à des changements dans les conditions d'illumination, comme le montrent les évaluations FERET [BBP01] (*cf.* section 1.7.2).

2.2.2 Les approches de projection statistique

Notons $h \times w$ la résolution initiale des images de la base d'apprentissage. La plupart du temps, chaque image est représentée par un vecteur de pixels de très grande dimension $n = hw$, obtenu par concaténation des lignes ou des colonnes de pixels de la matrice-image initiale. L'espace $\mathcal{I}$ contenant l'ensemble des vecteurs-images de visages est appelé *espace des images*. Ses dimensions sont très importantes, ce qui rend la classification difficile dans cet espace.

Heureusement, les images de visages partagent un certain nombre de propriétés structurelles communes. En effet, sous une pose frontale, les visages sont à peu près symétriques, et comportent un certain nombre de caractéristiques faciales dont on connaît les localisations approximatives. Donc, les visages ne sont pas distribués de manière aléatoire dans $\mathcal{I}$ et une grande partie des points de l'espace $\mathcal{I}$ des images ne peuvent pas correspondre à des visages. De plus, dans le cas des visages, l'information contenue dans $\mathcal{I}$ est souvent redondante : en effet, les valeurs de pixels voisins sont généralement très corrélées. Par conséquent, on peut considérer que généralement les visages appartiennent à un sous-espace $\mathcal{F}$ de $\mathcal{I}$, de dimension inférieure, appelé *espace des visages* [SM04]. Les méthodes de *projection statistique*, aussi appelées *méthodes des sous-espaces* ou *de réduction de dimensions* visent, dans un premier temps, à définir ce sous-espace et, dans un deuxième temps, à mettre en correspondance les visages à l'intérieur de ce sous-espace. Pour un état de l'art très détaillé de ces techniques, se référer à [SM04].

Le problème de la projection statistique peut être formulé de la manière suivante : connaissant une variable aléatoire $n-$dimensionnelle $\mathbf{x} = (x_1, \ldots, x_n)^T$, on recherche une autre représentation $\mathbf{s} = (s_1, \ldots, s_g)^T$, avec $g < n$, optimale au sens d'un critère fixé. Nous pouvons classer ces techniques en deux grandes familles : les *méthodes linéaires* et les *méthodes non linéaires*.

Les méthodes linéaires visent à définir une nouvelle base de l'espace original des données $\mathcal{I}$. Une fois les données projetées linéairement dans cette nouvelle base, on élimine les vecteurs de base les moins porteurs d'*information* (au sens d'un critère bien choisi), définissant ainsi la base d'un espace de dimension réduite $\mathcal{F}$. La dimension g intrinsèque du sous-espace $\mathcal{F}$ peut donc être fixée *a posteriori*. Les données transformées par les techniques linéaires sont des combinaisons

linéaires des données originales, soit :

$$\mathbf{s} = W^T\mathbf{x} \tag{2.1}$$

où W est la matrice de taille $n \times g$ de la transformation linéaire. La transformée inverse est :

$$\mathbf{x} = A^T\mathbf{s} \tag{2.2}$$

où A est la matrice de taille $g \times n$ de transformation inverse.

Le but des méthodes non linéaires est de fournir un espace $\mathcal{F}$ de représentation dans lequel les données sont projetées non linéairement. Cela permet de trouver des hypersurfaces de représentation (ou de séparation) capables de représenter et de classer des données dont les distributions sont plus complexes. Ces techniques, la plupart du temps itératives, nécessitent donc souvent un choix *a priori* de la dimension intrinsèque g de l'espace $\mathcal{F}$.

Les principales méthodes linéaires utilisées dans le contexte de la reconnaissance de visages sont : l'Analyse en Composantes Principales (ACP), l'Analyse Discriminante Linéaire (ADL), et l'Analyse en Composantes Indépendantes (ACI). L'ACP est notamment utilisée dans le cadre des méthodes des *eigenfaces*, des sous-espaces probabilistes et des sous-espaces Bayésiens, détaillées dans les trois sections 2.2.2.1 - 2.2.2.3 ci-après. La technique des *fisherfaces*, basée sur l'ADL, est brièvement passée en revue en section 2.2.2.4 (cette dernière sera détaillée d'une manière plus poussée en section 3.3.3.2, p. 65). Les approches basées sur l'ACI sont étudiées en section 2.2.2.5. Les techniques de projection statistique non linéaire les plus courantes, visant à représenter des espaces de données plus complexes que les techniques linéaires, sont présentées en section 2.2.2.6. Nous introduirons en section 2.2.2.7 une alternative à ces derniers modèles, souvent difficiles à estimer, en la construction d'un ensemble de sous-espaces linéaires.

2.2.2.1 La méthode des *eigenfaces*

L'Analyse en Composantes Principales L'Analyse en Composantes Principales (ACP), initialement introduite par Hotelling en 1933 [Hot33], est une technique d'analyse de données permettant de définir le sous-espace décrivant le mieux possible la distribution des données dans l'espace initial. On suppose les données multinormales et centrées. L'ACP vise donc à définir le sous-espace $\mathcal{F}$ de l'espace initial $\mathcal{I}$ tel que la dispersion des données, projetées orthogonalement dans $\mathcal{F}$, soit maximale (voir figure 2.1). Le sous-espace déterminé par l'ACP est appelé *sous-espace principal*, et est engendré par une base orthonormée d'*axes principaux*.

Soit $\{A_1, A_2, \ldots, A_N\}$ l'ensemble des données dont on dispose, sous la forme de vecteurs de $\mathbb{R}^n$. Les données sont centrées, c.-à-d. que l'on a : $\overline{A} = \frac{1}{N}\sum_{l=1}^{N} A_l = 0$ Notons W la matrice constituant une base orthonormée de $\mathcal{F}$. La projection orthogonale de A_l sur W est donnée par :

$$A'_l = W^T A_l \tag{2.3}$$

et par conséquent la matrice de dispersion des données projetées dans $\mathcal{F}$ peut s'écrire :

$$S'_T = W^T S_T W \tag{2.4}$$

où la matrice S_T est la matrice de dispersion totale des données initiales dans $\mathcal{I}$:

$$S_T = \frac{1}{N} A_l A_l^T \tag{2.5}$$

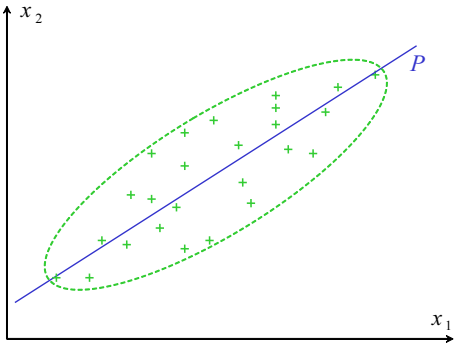


FIG. 2.1 – *Le sous-espace principal $\mathcal{F}$ est défini par l'axe principal P, qui donne la direction de $\mathcal{I} = \text{vect}(x_1,x_2)$ suivant laquelle la dispersion des données est maximale.*

Le critère à maximiser, basé sur la dispersion des données projetées, est donc :

$$J(W) \quad = \quad \left|W^T S_T W\right| \tag{2.6}$$

On peut montrer que les colonnes de W sont constituées des g vecteurs propres orthonormés de la matrice S_T (symétrique réelle) associés aux plus grandes valeurs propres [Loè55, Jol86]. La valeur propre associée à chaque vecteur propre est une mesure du pourcentage de variance expliqué par ce vecteur propre.

Le sous-espace principal vérifie deux propriétés majeures. La première est que, pour une taille g fixée, il minimise l'*Erreur Euclidienne de Reconstruction* moyenne ϵ, calculée selon :

$$\epsilon \quad = \quad \frac{1}{N}\sum_{l=1}^{N}\left\|A_l - \sum_{i=1}^{g}\left(W_i W_i^T A_l\right)\right\|_2 \tag{2.7}$$

La seconde est que l'ACP permet de *décorréler* les variables, en ce sens que les matrices de covariance des données $W_i^T A_l$ projetées sur chacun des axes discriminants W_i sont diagonales, pour tout i allant de 1 à g. Cette propriété assure la non-redondance des variables projetées, et donc le caractère optimal du sous-espace $\mathcal{F}$ choisi, pour une taille g fixée.

Les *eigenfaces* En 1987, Sirovitch et Kirby [SK87] ont donné une nouvelle impulsion à la reconnaissance automatique de visages, en montrant que l'ACP constitue un outil efficace pour la représentation des visages.

La technique que nous nommerons par la suite *méthode des eigenfaces* consiste à appliquer une ACP sur les vecteurs-images de visages. Si la dimension n des vecteurs-visages est très supérieure à leur nombre N ($n \gg N$), ce qui est généralement le cas pour les bases de visages, on peut utiliser une astuce courante [TP91]. Notons $A = [A_1, A_N]$ la matrice des observations centrées. Au lieu de calculer directement les éléments propres de la matrice $S_T = \frac{1}{N}AA^T$ (de très grande taille $n \times n$), on peut alternativement calculer le système propre de $A^T A \in \mathbb{R}^{N\times N}$, et en déduire le système propre de S_T par projection.

À chaque vecteur-image $A_l \in \mathbb{R}^n$ (supposé centré) est associée sa *signature* A'_l, définie par la projection de A_l sur W, selon (2.3). La classification s'effectue dans $\mathcal{F}$, le plus souvent par simple mesure de dissimilarité (*cf.* annexe D) entre signatures et une assignation au plus proche voisin.

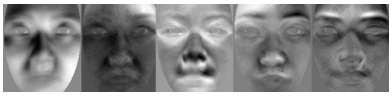


FIG. 2.2 – *Les cinq premières eigenfaces (associées aux plus grandes valeurs propres), calculées sur une sous-base de l'Asian Face Database PF01 (cf. annexe A), contenant 107 personnes et quatre images de visages par personne.*

Les colonnes de la matrice de projection W, dont des exemples sont fournis en figure 2.2, sont appelés *eigenpictures* dans [SK87, KS90], *eigenfaces* par Turk et Pentland [TP91] et *Most Expressive Features* dans [SW96]. Nous les désignerons sous le nom de *eigenfaces* dans la suite. Puisque ces *eigenfaces* sont choisies de manière à expliquer le mieux possible la distribution des images de visages, il est logique qu'elles soient représentatives, visuellement parlant, de la structure des visages. Néanmoins, on s'aperçoit qu'elles représentent également, et notamment les deux premières d'entre elles, des variations dans les conditions d'illumination. Ceci est imputable au fait que le critère de l'ACP (2.6) cherche à maximiser la variance totale, prenant ainsi en compte toutes les sources de variations, y compris le bruit[4].

Les choix de la dimensionnalité intrinsèque g du sous-espace principal, ainsi que de la mesure de dissimilarité utilisée, constituent deux enjeux majeurs.

Sélection des vecteurs propres à retenir Un point important et qui reste une voie de recherche est le choix du paramètre g. Celui-ci détermine la dimensionnalité intrinsèque de l'espace des visages. Pour déterminer la valeur optimale de g, on peut se baser sur l'étude du spectre des valeurs propres λ_i, par le biais d'un graphe de l'éboulis des valeurs propres (voir figure 2.3). Un algorithme naturel pour déterminer g est de chercher la valeur-charnière à partir de laquelle les valeurs propres (normalisées) sont très petites.

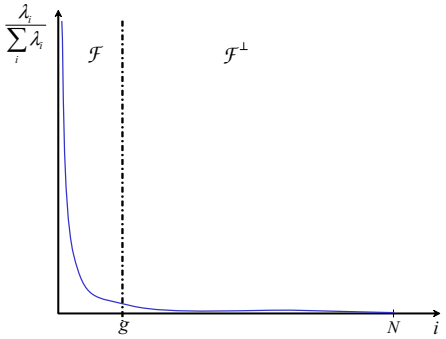


FIG. 2.3 – *Allure typique du graphe d'éboulis des valeurs propres associées à des eigenfaces. Les valeurs propres λ_i sont rangées par ordre décroissant. Le paramètre g est fixé à la valeur-charnière à partir de laquelle les valeurs propres normalisées $\frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i}$ sont très petites.*

Dans [TP91], g est défini de manière *heuristique* à partir de l'étude des valeurs propres, et donc non automatique. Moon et Phillips [MP98] préconisent d'éliminer 40% des derniers vecteurs propres, mais ce critère peut ne pas être optimal, selon le contenu de la base d'apprentissage.

4. On désigne par le terme *bruit* toute source de variation ne provenant pas d'un changement d'identité, mais de modifications dans la pose de la tête, les expressions faciales, les conditions de prise de vue, etc.

Kirby et Sirovitch [KS90] ont introduit un premier critère de sélection qui est par la suite devenu classique [SW96] et fut baptisé *énergie de dimension* [Kir00]. L'énergie de dimension du $i^{\text{ème}}$ vecteur propre est :

$$E_i = \frac{\sum_{j=i+1}^{n} \lambda_j}{\sum_{j=1}^{n} \lambda_j} \tag{2.8}$$

où λ_j est la valeur propre associée à la $j^{\text{ème}}$ *eigenfaces*. On peut montrer que $\sum_{j=i+1}^{n} \lambda_j$ est l'*erreur quadratique moyenne* engendrée par le rejet des $n-i$ derniers vecteurs propres ; le critère consiste à sélectionner les g premiers vecteurs propres tels que $E_{g-1} > \tau$, et $E_g < \tau$, où τ est un seuil fixé. Swets et Weng [SW96] préconisent l'utilisation de $\tau = 5\%$, et Kirby [Kir00] utilise $\tau = 10\%$. Kirby [Kir00] a introduit un critère d'*étirement*, défini comme le ratio $s_i = \frac{\lambda_i}{\lambda_1}$ entre la valeur propre de W_i et la plus grande valeur propre λ_1. Seuls les vecteurs propres dont les s_i sont supérieurs à un certain seuil τ sont retenus (souvent τ =1%). Ces deux méthodes de sélection offrent à peu près les mêmes performances en termes de taux de reconnaissance.

Il n'existe cependant aucune preuve que le fait de ne retenir que les vecteurs propres associés aux plus grandes valeurs propres garantisse une meilleure discrimination des visages selon leur identité. La figure 2.2 laisse même à penser que, loin d'encoder de l'information discriminante, les premiers vecteurs propres représenteraient du bruit. C'est pourquoi Moon et Phillips [MP98] préconisent d'éliminer le premier vecteur propre. Martinez et Kak [MK01] montrent que les résultats expérimentaux obtenus par la technique des *eigenfaces* peuvent être meilleurs si l'on ne prend pas en compte les trois premiers vecteurs propres, etc. Le nombre de vecteurs propres à rejeter est en fait très dépendant de la base d'apprentissage utilisée.

Choix de la mesure de dissimilarité la mieux adaptée La classification des signatures fournies par la technique des *eigenfaces* est généralement menée à l'aide d'une distance au plus proche voisin : le visage-requête est affecté à la classe d'appartenance dont la signature est la plus proche. Les distances les plus utilisées sont : la distance L_1 (Manhattan), L_2 (Euclidienne), du cosinus et de Mahalanobis [BSDG01] (*cf.* annexe D). Des combinaisons de ces quatre métriques de base, telles que les mesures de dissimilarité de Mahalanobis-L_1, -L_2 et -cosinus ont été proposées dans [YDB00]. Des variantes de ces distances usuelles, telles que les mesures de dissimilarité de Moon [MP98] et de Yambor [YDB00], ont également été introduites. Les résultats expérimentaux présentés dans [BBTD03], ainsi que nos expériences personnelles, montrent que la distance de Mahalanobis-cosinus est la plus performante d'entre elles.

2.2.2.2 Les sous-espaces probabilistes

L'ACP rejette purement et simplement les $n - g$ vecteurs propres associés aux plus faibles valeurs propres, ce qui peut engendrer une perte d'information. L'*Analyse en Composantes Principales Probabiliste* (ACPP), proposée par Roweis [Row97] et Tipping et Bishop [TB97] en 1997, est une extension de l'ACP traditionnelle prenant en compte les derniers vecteurs propres par le biais d'un modèle linéaire bruité. Tipping et Bishop [TB99b] ont montré que l'ACPP pouvait être reformulée comme la solution par maximum de vraisemblance d'un modèle spécifique de variable latente.

Un cas particulier de l'ACP Probabiliste est le modèle introduit par Moghaddam et Pentland [MP97], et basé sur la décomposition du sous-espace propre en deux espaces complémentaires : l'espace des visages $\mathcal{F}$ (déterminé par la méthode des *eigenfaces*), et son complémentaire $\mathcal{F}^{\perp}$, dont une base est définie par les vecteurs propres restants. L'espace des images $\mathcal{I}$ est

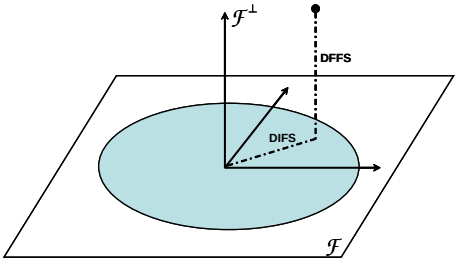


FIG. 2.4 – *La différence Δ est décomposée suivant la distance à l'intérieur de l'espace des visages (DIFS), et à l'extérieur de l'espace des visages (DFFS).*

donc décomposé en deux sous-espaces orthogonaux (voir figure 2.4). Pour mesurer la différence $\Delta = A_1 - A_2$ entre deux images de visages A_1 et A_2, on utilise donc deux distances. À l'intérieur de l'espace des visages $\mathcal{F}$ on considère la *Distance in Feature Space* (DIFS), et à l'extérieur de $\mathcal{F}$ on mesure la *Distance From Feature Space* (DFFS). Puisque les espaces $\mathcal{F}$ et $\mathcal{F}^\perp$ sont orthogonaux, on peut décomposer la densité de probabilité de la différence selon :

$$\mathbb{P}\left[\Delta/\Omega\right] \quad = \quad \mathbb{P}_{\mathcal{F}}\left[\Delta/\Omega\right]\mathbb{P}_{\mathcal{F}^\perp}\left[\Delta/\Omega\right] \tag{2.9}$$

où Ω est l'espace décrivant l'ensemble des images de visages possibles, décrit par le biais de la base d'apprentissage. On suppose vérifiées les hypothèses selon lesquelles les données sont distribuées selon une loi multinormale dans $\mathcal{F}$, et qu'elles sont localisées dans $\mathcal{F}$, à l'exception d'un bruit blanc Gaussien dans $\mathcal{F}^\perp$. On peut donc estimer $\mathbb{P}\left[\Delta/\Omega\right]$ par $\hat{\mathbb{P}}\left[\Delta/\Omega\right]$, selon [MP97] :

$$\hat{\mathbb{P}}\left[\Delta/\Omega\right] \quad = \quad \left[\frac{\exp\left(-1/2\sum_{i=1}^{g}\frac{(P_i^T\Delta)^2}{\lambda_i}\right)}{(2\Pi)^{g/2}\prod_{i=1}^{g}\lambda_i^{1/2}}\right]\cdot\left[\frac{\exp\left(-\frac{\epsilon^2(\Delta)}{2\rho}\right)}{(2\Pi\rho)^{(n-g)/2}}\right] \tag{2.10}$$

où $\epsilon(\Delta)$ est l'erreur moyenne de reconstruction de l'ACP (2.7), et ρ est simplement la moyenne des $n-g$ dernières valeurs propres. Dans la pratique, il est rare que l'on puisse calculer précisément ces valeurs propres, à cause d'un nombre d'exemples trop faible. Dans ce cas, Moghaddam et Pentland [MP97] préconisent d'adapter une fonction non linéaire sur la portion connue du spectre des valeurs propres, pour estimer les dernières valeurs propres. Lorsqu'une image-requête T doit être classée, on calcule sa différence avec chacune des images A_l de la base d'apprentissage et on décide de lui assigner la classe d'appartenance de l'exemple le plus proche selon (2.10). Si le ratio DFFS/DIFS dépasse un certain seuil, on peut décider que le visage-requête n'est pas représenté dans la base d'apprentissage.

2.2.2.3 La méthode des sous-espaces Bayésiens

On considère maintenant qu'il existe deux types de variations pour la différence $\Delta = A_1 - A_2$ entre deux images : les variations *intra-classe* Ω_I et les variations *inter-classe* Ω_E. Les variations intra-classe correspondent, pour une même personne, à des expressions faciales, des poses, etc. différentes, et sont modélisées à partir de l'ensemble des vues d'une même personne (dans la base d'apprentissage). Les variations inter-classe proviennent des différences structurelles entre visages différents. Une mesure de similarité $S(\Delta)$ entre les images A_1 et A_2 peut être définie par la probabilité que, compte tenu de la différence Δ observée, les deux images A_1 et A_2 appartiennent à la même classe :

$$S(\Delta) \quad = \quad \mathbb{P}\left[\Omega_I/\Delta\right] = \frac{\mathbb{P}\left[\Delta/\Omega_I\right]\mathbb{P}\left[\Omega_I\right]}{\mathbb{P}\left[\Delta/\Omega_I\right]\mathbb{P}\left[\Omega_I\right] + \mathbb{P}\left[\Delta/\Omega_E\right]\mathbb{P}\left[\Omega_E\right]} \tag{2.11}$$

Ainsi, on passe d'un problème de classification à k groupes à un problème de classification binaire. Les probabilités *a priori* $\mathbb{P}[\Delta/\Omega_I]$ et $\mathbb{P}[\Delta/\Omega_E]$ sont estimées après application d'une ACP dans chacun des sous-espaces Ω_I et Ω_E. Sous l'hypothèse que les données sont distribuées selon des lois multinormales dans Ω_I et Ω_E, on peut définir une expression analytique de la mesure de similarité $S(\Delta)$ [MJP00]. Lorsqu'un visage-requête T doit être classé, on calcule sa différence $\Delta_l = T - A_l$ à chacune des images A_l connues, ainsi que la mesure de similarité $S(\Delta_l)$ associée, et on assigne à T la classe d'appartenance de l'exemple A_l le plus proche.

On peut remarquer qu'il s'agit d'une variante *supervisée* de l'ACP, qui est elle-même non supervisée. Le terme *supervisé* signifie que chaque individu de la base d'apprentissage est étiqueté par sa classe d'appartenance : on suppose à l'origine un système de classes établi, donné *a priori*. Cette approche est cependant plus coûteuse que la technique des *eigenfaces*, et dépend de deux paramètres (les dimensions des deux sous-espaces principaux considérés), sans qu'aucune stratégie de choix de ces paramètres n'ait été proposée par ses auteurs.

2.2.2.4 La méthode des *fisherfaces*

La méthode des *fisherfaces* que nous décrivons dans cette section est sans doute la plus connue des approches utilisant l'Analyse Discriminante Linéaire (ADL) dans le contexte de la reconnaissance de visages. Il existe de nombreuses méthodes de reconnaissance de visages basées sur l'ADL. La plupart de ces techniques seront présentées dans le chapitre 3. À cette occasion, nous reviendrons plus en détail sur l'ADL (section 3.2) et la technique des *fisherfaces* (section 3.3.3.2).

L'Analyse Discriminante Linéaire L'Analyse Discriminante Linéaire (ADL) est une technique supervisée, basée sur la maximisation d'un critère de séparabilité [Fis36]. On dispose d'une base d'apprentissage Ω composée de N observations A_l de $\mathbb{R}^n$, chacune étant affecté à l'une des k classes. Le nuage de points Ω est donc partagé en k sous-nuages $\Omega_1, \Omega_2, \ldots, \Omega_k$, chacun de ces sous-nuages correspondant à une classe. Il s'agit de trouver le sous-espace maximisant par projection les dissimilarités entre classes, tout en réduisant au maximum les variations à l'intérieur des classes (expliquant la majeure partie du bruit). Le sous-espace linéaire $\mathcal{F}$ est engendré par la matrice W maximisant le critère de Fisher suivant [BHK97] :

$$J(W) = \frac{\left|W^T S_b W\right|}{|W^T S_w W|}$$

où

$$S_w = \frac{1}{N}\sum_{j=1}^{k}\sum_{A_l\in\Omega_j}(A_l - \overline{A_j})(A_l - \overline{A_j})^T$$

$$S_b = \frac{1}{N}\sum_{j=1}^{k} N_j(\overline{A_j} - \overline{A})(\overline{A_j} - \overline{A})^T$$

où Ω_j désigne l'ensemble des vues de la $j^{\text{ème}}$ personne, $\overline{A_j}$ est la moyenne des observations issues de Ω_j, et $\overline{A} = \frac{1}{N}\sum_{j=1}^{k} N_j\overline{A_j}$ est la moyenne de l'ensemble des observations issues de Ω. Les données sont supposées centrées, c.-à-d. $\overline{A} = 0$. La matrice S_w, qui est la moyenne des variances à l'intérieur des classes, est appelée *matrice de variance intra-classe*, tandis que la matrice S_b, mesurant la variance de la moyenne des classes, est appelée *matrice de variance inter-classe*. Sous l'hypothèse que la matrice S_w est inversible, les colonnes de la matrice W sont constituées des vecteurs propres de la matrice $S_w^{-1}S_b$, associés aux plus grandes valeurs propres.

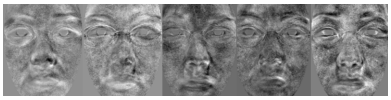


FIG. 2.5 – *Les cinq première fisherfaces (associées aux plus grandes valeurs propres), calculées sur une sous-base de l'Asian Face Database PF01 (cf. annexe A), contenant 107 personnes et quatre images de visages par personne.*

Tandis que l'ACP cherche à représenter au mieux les nuages de points (correspondant aux classes), l'ADL cherche la direction permettant de mieux les séparer, comme l'illustre la figure 3.2, donnée en p. 59. Si les données sont multinormales et homoscédastiques (les matrices de variance intra-classe des différentes classes sont égales), alors l'ADL définit un classifieur Bayésien optimal (voir section 3.2.3).

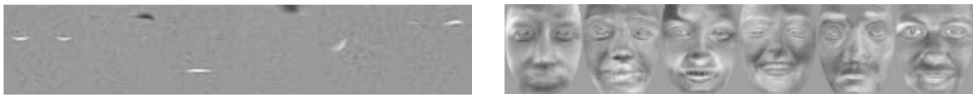
Les *fisherfaces* Dans le contexte de la reconnaissance de visages, les données sont le plus souvent *sous-représentées*, au sens où la taille n des vecteurs-images de la base d'apprentissage est très supérieure à leur nombre ($n \gg N$). La matrice S_w est alors non inversible et on ne peut pas déterminer directement W : c'est le *problème de la singularité*, détaillé en section 3.3.2. Une solution à ce problème, proposée par Swets et Weng [SW96] et Belhumeur *et al.* [BHK97] est d'effectuer une ACP en amont de l'ADL. Cette phase préliminaire permet de réduire la dimensionnalité du problème, de manière à ce que la nouvelle matrice de variance intra-classe soit inversible, et ceci en conservant au maximum la forme de la distribution initiale des données.

Les colonnes de la matrice de projection W, dont des exemples sont fournis en figure 2.5, sont appelés *fisherfaces* par Belhumeur *et al.* [BHK97] et *Most Discriminant Eigenfeatures* dans [SW96]. Nous les désignerons sous le nom de *fisherfaces* dans la suite. Visuellement parlant, les *fisherfaces* sont moins représentatives de la structure des visages que les *eigenfaces* puisque le critère à maximiser n'est plus lié à la qualité de représentation, mais à la séparabilité. Bien que construites depuis la même base que les *eigenfaces* données en figure 2.2, on constate que les *fisherfaces* données en figure 2.5 semblent moins représentatives des variations d'illuminations. En effet, celles-ci sont en grande partie expliquées par la variance intra-classe, que le critère de Fisher tend à minimiser.

Choix de la mesure de dissimilarité la mieux adaptée Tout comme pour les *eigenfaces*, les signatures obtenues par la technique des *fisherfaces* sont généralement comparées à l'aide d'une distance au plus proche voisin. Comme nous le montrons en section E.1 (p. 187) de l'annexe E, les valeurs propres associées aux vecteurs propres de $S_w^{-1}S_b$ sont liées au pouvoir discriminant de ces derniers. Plus la valeur propre est grande, plus le vecteur propre associé permet de mieux séparer les différentes classes. Les valeurs propres peuvent donc être incorporées dans la mesure de dissimilarité utilisée. Par exemple Zhao propose dans [Zha99] une distance Euclidienne pondérée par les valeurs propres (*cf.* annexe D). Celle-ci donne de meilleurs résultats de reconnaissance que la plupart des métriques usuelles.

2.2.2.5 Les approches basées sur l'Analyse en Composantes Indépendantes

L'Analyse en Composantes Indépendantes La technique d'Analyse en Composantes Indépendantes (ACI) [Com94, Hyv99] est habituellement utilisée pour la séparation de sources [JH91], mais est très prisée dans de nombreux domaines. Celle-ci vise à trouver le sous-espace linéaire le plus représentatif de la distribution initiale des données. Si l'on ne fait aucune hypothèse



(a) **(b)**

FIG. 2.6 – *Extrait de [DBBB03]. Les six premières colonnes de la matrice W, obtenues (a) grâce à l'architecture I et (b) avec l'architecture II.*

concernant cette distribution, il n'y a aucune raison de s'arrêter à des statistiques d'ordre deux. L'ACI vise à représenter également les statistiques d'ordre supérieur ; autrement dit, au lieu de simplement *décorréler* les données comme le fait l'ACP, l'ACI cherche à les rendre *statistiquement indépendantes*. L'ACI peut donc être vue comme une extension non paramétrique de l'ACP. Mais, à la différence de l'ACP, l'ACI n'engendre pas de réduction de dimension. En pratique, l'ACI est la plupart du temps utilisée après une phase préliminaire d'ACP, visant à réduire les dimensions du problème en décorrélant les données.

Appliquer l'ACI sur un vecteur aléatoire $n-$dimensionnel $\mathbf{x}$ consiste à estimer le modèle :

$$\mathbf{x} = A^T\mathbf{s} \tag{2.12}$$

où les variables s_i du vecteur $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ sont supposées statistiquement indépendantes et où A est une *matrice de mélange* de taille $g \times n$.

L'ACI consiste à optimiser une fonction objectif : la *fonction de contraste*. Celle-ci peut être basée sur le maximum de vraisemblance, par le biais du calcul de la divergence de Kullback-Leibler par exemple (voir annexe D), ou sur la non-normalité des variables. En effet, la théorie de la « poursuite de la projection » [Hub85] dit que chercher l'indépendance statistique revient à chercher la non-normalité. Le sous-espace de projection est donc choisi de manière à ce que les données projetées dans ce sous-espace soient non-Gaussiennes. On utilise souvent pour cela des fonctions de contraste basées sur une approximation de la néguentropie[5]. Il existe de nombreux algorithmes de mise en œuvre de l'ACI, dont les plus connus sont INFOMAX [BS95], JADE [Car99], et FASTICA [HO97].

L'ACI pour la reconnaissance automatique de visages Bartlett *et al.* ont proposé, pour la reconnaissance de visages, deux algorithmes de mise en œuvre fondamentalement différents [BMS02]. Tous deux reposent sur une étape préliminaire d'ACP. Le premier algorithme : « architecture I » vise à obtenir des vecteurs de base (les colonnes de la matrice W) qui soient statistiquement indépendants deux à deux. L'algorithme « architecture II », quant à lui, cherche à rendre les coefficients de projection (variables) mutuellement statistiquement indépendants. Les vecteurs de projection de l'ACI, à l'instar des *eigenfaces* pour l'ACP et des *fisherfaces* pour l'ADL, sont illustrés en figure 2.6. L'ACI n'engendre pas de réduction de dimension. Par conséquent, les vecteurs propres sont en même nombre que la dimensionnalité choisie pour l'ACP préliminaire. On peut remarquer que l'architecture I fournit des vecteurs de base expliquant essentiellement des propriétés locales, tandis que les vecteurs issus de l'architecture II semblent fournir plus d'information sur la globalité du visage. La classification est effectuée à l'aide d'une mesure de similarité au plus proche voisin. Selon les expérimentations de Bartlett *et al.* [BMS02], ainsi que celles de Delac *et al.* [DGG05], la distance du cosinus est la plus adaptée à l'ACI.

5. Sachant qu'une variable gaussienne X^* a une plus grande entropie (au sens de sa définition probabiliste) H que n'importe quelle variable aléatoire X de même variance, une mesure de non Gaussiannité de X est donnée par sa néguentropie $N(X) = H(X^*) - H(X)$.

Dans la littérature, les résultats de comparaison des performances de l'ACI et de l'ACP sont contradictoires. Les résultats fournis par Bartlett *et al.* [BMS02] (sur une sous-base de FERET) mettent en évidence des performances équivalentes pour les deux architectures, et supérieures à celles des *eigenfaces*. Les résultats obtenus par Delac *et al.* [DGG05], ainsi que par Draper *et al.* [DBBB03] (sur des sous-bases de FERET) montrent une différence très significative dans les performances des deux algorithmes, à l'avantage de l'architecture II. Selon ces deux références, seul le modèle issu de la seconde architecture serait plus efficace que l'ACP. Néanmoins, la plupart des résultats expérimentaux laissent à penser que l'ACI n'apporte pas d'amélioration significative sur l'ACP [Mog02], voire qu'elle engendre une dégradation des performances [BDBS02, Yan02].

2.2.2.6 Les sous-espaces non linéaires

Les techniques basées sur la construction d'un sous-espace non linéaire ont été introduites dans le contexte de la reconnaissance de visages dans le but de parvenir à représenter plus précisément les données, lorsque la distribution de celles-ci est complexe. Typiquement, si les conditions d'illumination changent drastiquement, les techniques non linéaires sont réputées plus performantes que les méthodes linéaires.

Les Courbes Principales La technique des *Courbes Principales* [HS89] consiste en un modèle de régression non linéaire des données. L'un des algorithmes de mise en œuvre les plus simples, et dans la plupart des cas équivalent aux Courbes Principales[6], est basé sur la construction d'une ACP non linéaire par le biais d'un réseau de neurones multicouches auto-associatif [CF90, Kra91]. L'architecture de ce réseau est illustrée en figure 2.7. La sortie désirée du réseau de neurone est

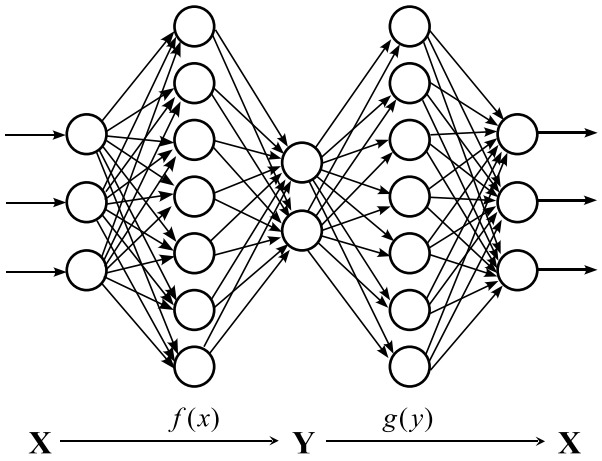


FIG. 2.7 – *Architecture du réseau de neurones auto-associatif permettant de déterminer les composantes principales non linéaires.*

égale au signal d'entrée. L'une des couches cachées, notée Y, comporte un nombre g de neurones fixé (qui est la dimension du sous-espace désiré). On cherche donc le sous-espace de dimension g fixée qui fournisse la meilleure reconstruction des données initiales. La partie du réseau de neurones comprise entre les données d'entrée et la couche Y correspond à une projection non linéaire des données $f(x)$ sous la forme d'une somme pondérée de fonctions sigmoïdes ; puis

6. Sous la condition que la fonction de projection non linéaire soit lisse et différentiable.

la portion du réseau comprise entre la couche Y et la sortie approxime la fonction inverse de reconstruction $h(y)$. À chaque visage est associé le vecteur des sorties de la couche cachée Y qui lui est associé, et constitue sa signature.

Cette technique permet de trouver un sous-espace optimal au sens de la reconstruction des données initiales et est donc, tout comme l'ACP, plus adapté à la compression des données qu'à leur classification. Un autre désavantage de cet algorithme est qu'il nécessite de fixer *a priori* la dimensionnalité intrinsèque g de l'espace des visages.

Utilisation d'une fonction de noyau L'utilisation d'une *fonction de noyau* (*cf.* section 3.5.1, p. 80), conjointement avec une méthode de projection statistique linéaire, est une astuce visant à rendre cette méthode non linéaire. La technique ainsi définie est conçue pour les cas où une technique de projection statistique linéaire ne suffit pas à séparer correctement les classes dans l'espace initial des données. Dans un premier temps, on projette les données dans un espace $\mathcal{K}$ de plus grande dimension et appelé *espace de linéarisation* [ABR64]. Dans un second temps, on applique dans $\mathcal{K}$ une technique de projection statistique linéaire. L'hyperplan ainsi obtenu peut être décrit dans l'espace initial des données (par projection) comme un sous-espace non linéaire. Notons qu'il n'est pas nécessaire de définir précisément $\mathcal{K}$ mais qu'il suffit de choisir une fonction de noyau adaptée au problème (*cf.* section 3.5.1, p. 80). En un certain sens, l'utilisation d'une fonction de noyau a donc rendu non linéaire une technique initialement linéaire. Ce principe est illustré en figure 3.6, p. 80. Des versions à noyau ont été proposées pour l'ACP [SSM99, Yan02], l'ADL [MRW+99], et l'ACI [BJ02]. Le choix du type de fonction de noyau à utiliser, ainsi que de ses paramètres, reste un problème difficile [GAP+02]. En section 3.5, nous étudierons en détails les différentes techniques d'ADL à noyau.

2.2.2.7 Extensions à plusieurs espaces de projection

Quand on dispose pour l'apprentissage de nombreuses vues correspondant à différents visages sous des conditions de prise de vue (pose de la tête, conditions d'illumination) variables, plusieurs stratégies sont possibles. La première approche consiste à utiliser toutes ces vues pour construire un unique sous-espace qui soit représentatif de l'ensemble des variations de la base d'apprentissage. De cette manière, le sous-espace décrit non seulement les identités des visages, mais aussi l'ensemble des variations dans les conditions de prise de vue. C'est par exemple la solution retenue par Murase et Nayar pour la reconnaissance d'objets en 3D [MN95]. On parle de *représentation paramétrique* des visages. Dans le cas de techniques linéaires, on suppose que les données reposent dans un unique sous-espace linéaire, et ce malgré leur complexité. Afin de capturer ces données complexes, nous avons vu ci-avant que des techniques non linéaires ont été introduites, avec plus ou moins de succès. Une approche alternative consiste à modéliser la distribution complexe des données par un ensemble de sous-modèles linéaires. Les données sont partitionnées en différents *clusters*[7]. On construit un espace de projection linéaire spécifique à chacun de ces *clusters*. On cherche à obtenir une meilleure représentation des données qu'avec un unique sous-espace linéaire, tout en évitant la complexité numérique inhérente à la détermination d'un sous-espace non linéaire. Il existe différentes manières de composer les *clusters* : on peut partitionner les données en fonction des conditions de prise de vue, et ceci de manière supervisée ou non, ou bien alors constituer un *cluster* par classe (identité).

7. On désignera sous le terme *cluster* un ensemble fini non ordonné de points d'un espace généralement multi-dimensionnel.

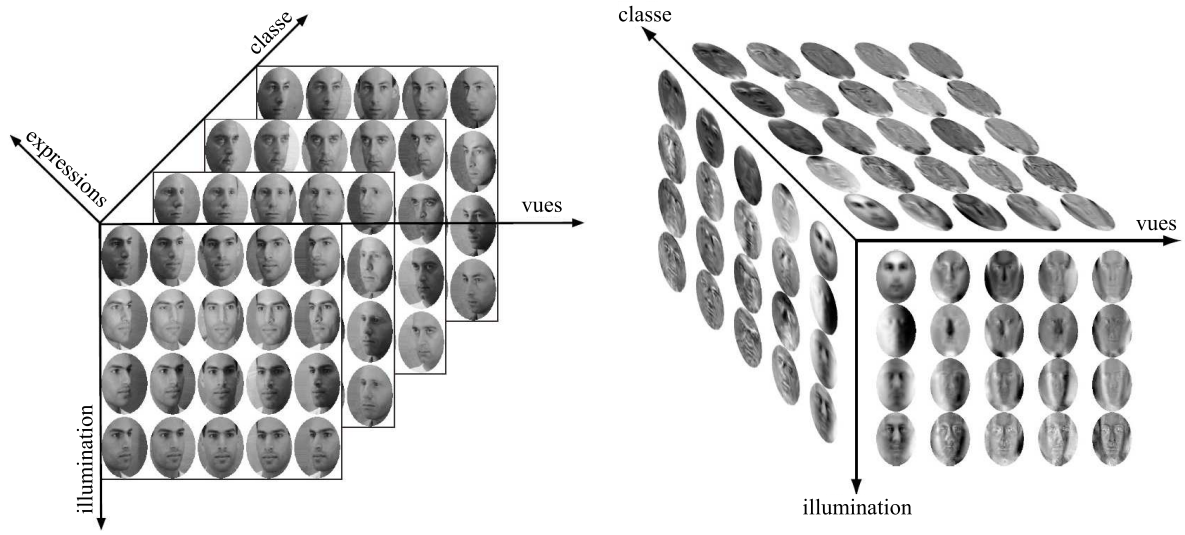


FIG. 2.8 – *Adapté de [VT02]. La technique de Vasilescu et Terzopoulos : (a) les données sont rangées sous la forme d'un tenseur, les quatre dimensions de celui-ci représentent la classe d'appartenance, la vue considérée (sous la forme d'un vecteur de pixels), les conditions d'illumination et l'expression faciale. Ici, seul le sous-tenseur correspondant à une expression faciale neutre est montré. (b) Le résultat de l'algorithme de Décomposition en Valeurs Singulières à n modes.*

Penchons-nous tout d'abord sur l'approche consistant à construire un sous-espace pour chaque type de prise de vue.

L'approche de Pentland *et al.* [PMS94] consiste à construire un sous-espace pour chaque couple d'orientation et d'échelle du visage dans l'image. Lorsqu'un visage-requête doit être reconnu, on commence par le projeter dans chacun de ces sous-espaces, et à sélectionner celui dont il est le plus proche. Puis, c'est dans ce sous-espace qu'est menée la phase de reconnaissance, de la même manière que pour les *eigenfaces*. La qualité de reconstruction obtenue à l'aide de cette méthode est meilleure que pour la technique paramétrique [SM04].

Vasilescu et Terzopoulos [VT02] ont plus récemment généralisé cette technique par l'utilisation de *tenseurs*. Un tenseur peut être vu comme une extension multidimensionnelle d'un vecteur (tenseur à une dimension) ou d'une matrice (tenseur à deux dimensions). Chaque vue est représentée par un vecteur de taille $n = hw$ pixels. Les données sont stockées sous la forme d'un tenseur, comme illustré en figure 2.8-a. L'algorithme proposé, appelé *Décomposition en Valeurs Singulières à n modes*, permet de décomposer le tenseur en un ensemble de composantes principales dans chacune des directions. Les vecteurs de base ainsi obtenus sont donnés en figure 2.8-b. Lorsqu'un visage-requête se présente, on commence par calculer pour chaque position dans le tenseur (c.-à-d. pour chaque pose, expression, etc.), ses coordonnées dans la base correspondante. On obtient ainsi un ensemble de vecteurs de coefficients. On choisit d'assigner au visage-requête l'identité de l'exemple le plus proche en moyenne sur toutes les positions dans le tenseur.

Kim *et al.* [KKB02], ainsi que Turaga et Chen [TC02], ont proposé des modèles de mélanges d'*eigenfaces* construits automatiquement par le biais d'un algorithme *Expectation Maximization* (EM) [Moo96]. Tandis que Kim *et al.* caractérisent chaque visage par ses coordonnées dans la base des *eigenfaces* la plus proche, Turaga et Chen choisissent comme signature le vecteur de coordonnées associé à la base des vecteurs propres minimisant l'erreur de reconstruction.

Les trois approches détaillées dans ce paragraphe sont plus performantes que la représentation paramétrique si la base contient de nombreuses variations d'apparence des visages. Mais ces approches nécessitent beaucoup d'images par personne, nécessairement dans des conditions de prise de vue différentes et souvent identifiées (sauf pour les techniques de mélanges d'*eigenfaces*). À titre d'exemple, Vasilescu et Terzopoulos [VT02] disposent de 256 vues par personne. Or, dans la pratique, il est très rare de disposer d'un tel échantillon étiqueté, ce qui limite le champ des applications.

Focalisons-nous maintenant sur les techniques reposant sur la construction d'un sous-espace par classe. Dans ce contexte, il n'est pas nécessaire que les images soient étiquetées par leurs conditions de prise de vue. Généralement, il suffit de disposer de moins d'images par personne qu'avec les trois techniques détaillées ci-avant. Aussi le champ des applications est-il plus large avec cette technique. Les bases de visages nécessaires à la construction du modèle peuvent être issues de vidéos segmentées en séquences d'images du même visage (à l'aide d'un algorithme de détection/suivi de visages).

Parmi les méthodes les plus intuitives, on peut citer la technique de Torres *et al.* [TLV00], permettant de comparer un visage à un ensemble de vues d'une même personne. Les sous-espaces sont définis en appliquant une ACP par classe (à la manière des *eigenfaces*). L'image du visage-requête est projetée dans chacun de ces sous-espaces, puis reconstruite. On assigne au visage à reconnaître l'identité associée au sous-espace donnant la plus faible erreur de reconstruction. Si l'on dispose non pas d'une unique image-requête, mais d'un ensemble de vues-requêtes de la même personne, alors on applique cette comparaison à toutes les vues-requêtes. La décision globale est prise par le biais d'un vote à la majorité.

Yamaguchi *et al.* a proposé en 1998 [YFM98] la technique dite des *Sous-Espaces Mutuels*, permettant de comparer directement des ensembles d'images. Lorsqu'un ensemble d'images-requêtes (contenant toutes le même visage) doit être classé, on construit son sous-espace principal par le biais d'une ACP (à la manière des *eigenfaces*), puis on calcule la distance entre ce sous-espace et les sous-espaces des personnes connues (préalablement construits) au sens des *angles principaux*. L'angle principal entre deux sous-espaces est défini comme étant l'angle minimum entre deux points des sous-espaces. Cette mesure ne prend en compte que la distance d'angle entre les deux points les plus proches et ceci quelles que soient les distributions des sous-espaces (on néglige notamment les centroïdes et les directions principales des sous-espaces). Elle peut par conséquent engendrer une perte d'information discriminante. Cette technique a été étendue à des sous-espaces non linéaires par Wolf *et al.* [WS03].

Une approche probabiliste permettant de mesurer la similarité entre espaces a été proposée dans [SFD02]. On cherche la classe Ω_j dont la densité de probabilité p_j est la plus proche de la distribution p de l'ensemble d'images à reconnaître, ces distributions étant supposées Gaussiennes. La mesure de dissimilarité considérée est la *divergence de Kullback-Leibler* (*cf.* annexe D). Les résultats expérimentaux montrent que cette technique probabiliste est plus performante que la technique des sous-espaces mutuels.

En 1999, Tipping et Bishop étendent la technique d'ACP Probabiliste présentée en section 2.2.2.2 à des sous-espaces probabilistes locaux [TB99a], dont les paramètres sont appris *via* un algorithme EM. L'avantage principal de cette technique est que l'estimation des distributions des classes permet de calculer les probabilités *a posteriori* d'appartenance à chacune des classes. Elle est utilisée avec succès pour la reconnaissance d'écriture manuscrite.

Récemment, Bouveyron *et al.* [BGS05] ont introduit une modélisation similaire à celle de Tipping et Bishop [TB99a], mais qui repose sur des sous-espaces de dimensions intrinsèques différentes (et déterminées à l'aide de l'étude des graphes d'éboulis des valeurs propres) et une

régularisation des matrices de variance des classes reposant sur l'hypothèse que les classes sont hypersphériques à la fois dans leur espace de projection et son supplémentaire. Cette approche est particulièrement adaptée à la modélisation des données de grandes dimensions et a été appliqué à la détection d'objets spécifiques (motocyclettes) dans des images, mais à notre connaissance, n'a jamais été utilisée dans le contexte de la reconnaissance de visages.

2.2.3 Les Modèles Actifs d'Apparence

Les Modèles Actifs d'Apparence (MAA) [CET01] constituent un outil d'extraction de signatures caractérisant à la fois la forme et la texture des visages. La base d'apprentissage est annotée très précisément (à la main) par un nombre de points caractéristiques important (on utilise couramment 122 points) modélisant la forme des visages (position des yeux, coins de la bouche, etc.). Les MAA ont été utilisés pour la première fois dans le contexte de la reconnaissance de visages en 1995 [LTC95]. Le processus de classification, détaillé ci-après, est illustré en figure 2.9. Chaque

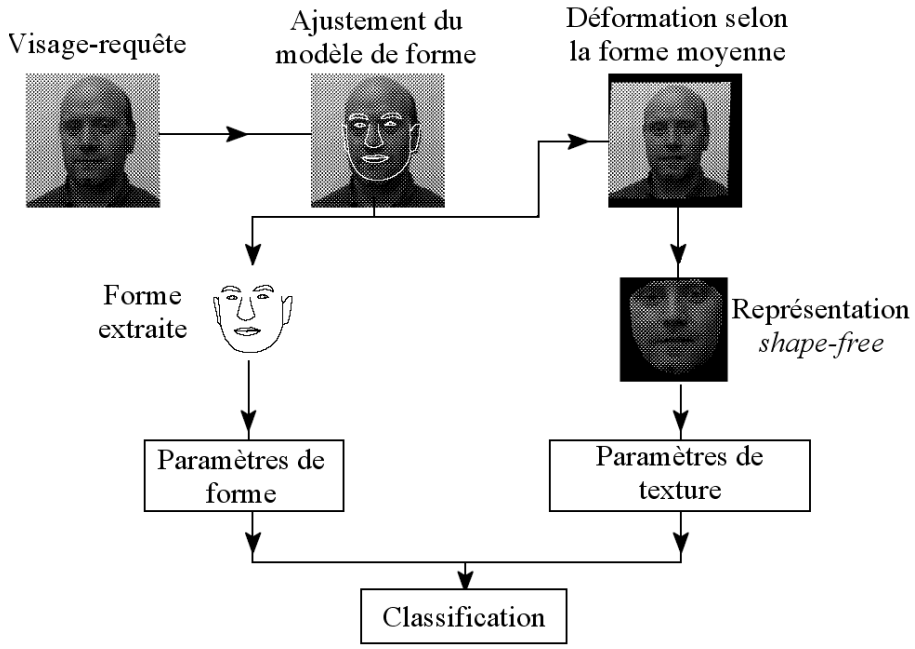


FIG. 2.9 – *Processus de reconnaissance de visages basé sur les Modèles Actifs d'Apparence.*

exemple est représenté par un *vecteur de forme* x contenant ses caractéristiques. À partir des vecteurs de forme de la base d'apprentissage, on applique une ACP, afin d'extraire un ensemble de modes de variations principaux de la forme. Ceux-ci sont stockés dans la matrice orthonormée P_f. Les vecteurs de forme x peuvent ensuite être estimés par le vecteur $\tilde{x}$ suivant :

$$\tilde{x} \quad = \quad \overline{x} + P_f b_f \tag{2.13}$$

où $\overline{x}$ est la forme moyenne de la base d'apprentissage et le vecteur $b_f = P_f^T x$ est la projection de x sur P_f, et constitue le vecteur des *paramètres de forme*. Les textures des visages (valeurs des pixels en niveaux de gris) sont alors normalisées au sens de leur forme. On obtient une représentation des visages dite *shape-free*. Les textures sont déformées selon la forme moyenne (à l'aide d'un algorithme de triangulation par exemple). Pour chaque visage, on obtient ainsi un vecteur de texture g, indépendant de la forme du visage. Une ACP est appliquée sur ces vecteurs

de texture ; le vecteur g peut donc être approximé par :

$$\tilde{g} = \overline{g} + P_g b_g \tag{2.14}$$

où b_g est le vecteur de *paramètres de texture* associé au visage. Chaque image est donc caractérisée par ses vecteurs de forme et de texture b_f et b_g, qui sont corrélés. Afin de les décorréler, on concatène les vecteurs b_f et b_g de chaque exemple de la base d'apprentissage, et l'on applique une ACP sur les vecteurs concaténés ainsi obtenus. On obtient alors le modèle combiné suivant :

$$\tilde{x} = \overline{x} + Q_f c \tag{2.15}$$
$$\tilde{g} = \overline{g} + Q_g c \tag{2.16}$$

où c est le *vecteur d'apparence* contrôlant à la fois la forme et la texture du modèle, et Q_f et Q_g sont respectivement les matrices de projection de c dans les espaces de variations de forme et de texture.

Pour un vecteur de paramètres c fixé, on peut synthétiser l'image de visage associée. Pour cela, on génère le visage *shape-free* correspondant, puis on procède à sa déformation en utilisant les points de contrôle du vecteur $\tilde{x}$. Lorsqu'un visage-requête doit être reconnu, le but est de déterminer le paramètre c optimal. Il s'agit de la valeur de c minimisant l'erreur δI entre l'image originale et l'image synthétisée. Pour cela, on applique une procédure d'optimisation itérative complexe dépendant d'un nombre de paramètres de l'ordre de 80 à 100. Afin de rendre ce processus plus rapide, Edwards *et al.* [ETC98] ont proposé d'introduire de la connaissance *a priori* dans l'optimisation. Cette information porte sur les relations entre l'erreur et l'ajustement des paramètres, et est apprise hors-ligne selon un modèle linéaire. Depuis, d'autres algorithmes de recherche itérative ont été introduits [CET01]. Ce processus itératif nous permet d'obtenir la valeur optimale du vecteur c, que l'on considère comme étant la signature du visage associé. La classification peut se faire à l'aide d'une distance de Mahalanobis [KAG03] entre les vecteurs c.

Cette technique est très utilisée dans le contexte de l'analyse de visages, non seulement pour leur reconnaissance [LTC95, ECT98, KAG03], mais aussi pour la détection et le suivi de visages et/ou d'éléments faciaux [Ahl01, CC04] (voir section 1.6.2), la modélisation de la forme 3D des visages [RPG99] et la reconnaissance d'expression faciale [DAD03]. Mais les MAA présentent le désavantage de reposer sur une procédure d'optimisation coûteuse, instable et dépendant de nombreux paramètres.

2.2.4 Les Réseaux de Neurones

Les réseaux de neurones artificiels ont été appliqués à la reconnaissance de visages, à la fois pour l'extraction de signatures et la classification de celles-ci.

Extraction de signature Nous avons présenté en section 2.2.2.6 l'utilisation de réseaux de neurones auto-associatifs pour l'extraction de signatures non linéaires.

Lawrence *et al.* [LGTB97] ont introduit une autre technique d'extraction de signatures, basée sur l'utilisation de Cartes Auto-Organisatrices introduites par Kohonen [Koh89], couramment appelées *cartes de Kohonen*. Celles-ci permettent d'organiser des données de grandes dimensions de manière non supervisée, en effectuant simultanément la projection dans la carte et le *clustering*[8] des données. Au contraire de la plupart des techniques de *clustering*, les cartes de Kohonen préservent la topologie des classes : la similarité entre les données d'entrée est préservée en sortie.

8. Agencement des données en *clusters* (voir note de bas de page en p. 36).

Classification L'une des premières approches de classification des visages par réseaux de neurones repose sur le système appelé Wilkie, Aleksander and Stonham's Recognition Device (WISARD) [Sto84]. Un réseau de neurones à une seule couche est construit pour chacune des classes. Le système nécessite pour son apprentissage de nombreuses vues d'une même personne, avec des variations dans les conditions d'illumination, l'expression faciale, etc. Un visage-requête se voit assigner l'identité du réseau de neurones qui produit la plus forte réponse.

Cotrell et Fleming [CF90] proposent d'effectuer la classification à l'aide d'un réseau Perceptron Multi-Couches (PMC), après extraction des composantes principales non linéaires par réseaux de neurones auto-associatifs (voir section 2.2.2.6). Dans [LGTB97], Lawrence *et al.* choisissent de classer les signatures (extraites par cartes de Kohonen) à l'aide d'un Réseau de Neurones Convolutionnel. Ce type de réseau de neurones est partiellement invariant à des transformations globales telles que la translation, la rotation et les changements d'échelles. Les résultats expérimentaux ont montré la supériorité des réseaux de neurones convolutionnels sur les réseaux de PMC, et une légère amélioration par rapport à la technique des *eigenfaces*.

Dans [LKL97], Lin *et al.* ont suggéré l'utilisation d'un Réseau de Neurones Probabiliste Décisionnel (alliant les avantages des approches statistiques et des réseaux de neurones). Il a été montré que les performances de cette solution sont comparables à la méthode de Lawrence *et al.* présentée ci-avant, tout en étant beaucoup moins coûteuse en termes de temps de calcul.

Les Réseaux de Fonctions à Base Radiale (RFBR) ont également été utilisés dans le contexte de la reconnaissance de visages. Les RFBR constituent une famille particulière de réseaux multicouches supervisés, comportant une couche cachée et une couche de sortie. Chaque neurone de la couche cachée implémente une *Fonction à Base Radiale* (voir section 5.4.2) définissant une hypersurface d'activation localisée autour d'un centre. Les valeurs de sortie sont des combinaisons linéaires des valeurs de ces fonctions. Les RFBR sont ainsi capables d'approximer n'importe quelle fonction, par combinaison linéaire d'un ensemble d'applications localisées (ce qui, par exemple, dans le cadre de Gaussiennes, revient à un modèle de mélange de Gaussiennes). Plus de détails concernant les RFBR sont fournis en section 5.4.2. Dans [WJHT04], Wang *et al.* proposent d'appliquer sur les visages une variante à noyau de l'*algorithme des K-moyennes* [JD88] afin d'initialiser les paramètres du RFBR. Les taux de classification obtenus sur la base ORL ne montrent pas d'amélioration par rapport aux techniques usuelles de projection statistique. Thomas *et al.* [TFV98] ont proposé d'utiliser un RFBR pour la classification des signatures extraites à l'aide de la technique des *eigenfaces*. Les mêmes auteurs ont montré dans [FTV99] que l'utilisation d'un RFBR en aval de l'ACP donne des résultats équivalent à l'utilisation d'une ADL (ce qui revient à appliquer l'algorithme des *fisherfaces*). Plus récemment, Er *et al.* [EWLT02] ont montré l'efficacité des Réseaux de Fonctions à Base Radiale (RFBR) pour la classification de signatures issues de la méthode des *fisherfaces*. Leur technique est néanmoins coûteuse en termes de construction du modèle, puisqu'elle nécessite la mise en œuvre d'une ACP, suivie d'une ADL, puis d'une initialisation itérative des paramètres du RFBR , et enfin de l'apprentissage de celui-ci. Plus de détails concernant la méthode d'initialisation utilisée seront donnés en section 5.4.2.

2.2.5 Les Machines à Vecteurs de Support

La technique des Machines à Vecteurs de Support (très connue sous son sigle anglais SVM) a été proposée en 1995 par Vapnik [Vap95]. Il s'agit d'une méthode de classification basée sur le concept de *minimisation du risque structurel*. Initialement, les SVM sont définis pour un problème binaire où les deux classes sont linéairement séparables. On peut néanmoins étendre leur définition aux cas non linéaire (par l'introduction d'une fonction de noyau) et non séparable (par relâchement des contraintes) ; enfin, il existe de nombreuses solutions pour l'extension au

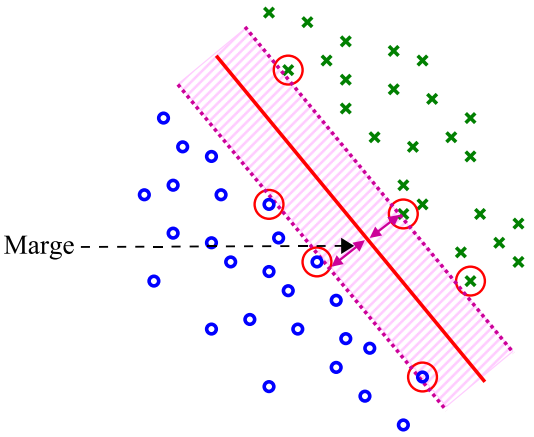


FIG. 2.10 – *Cas de deux classes linéairement séparables. L'hyperplan déterminé par la SVM, maximisant la marge, permet de séparer les deux classes de manière optimale. Les vecteurs de support sont entourés.*

cas multi-classes.

La technique de minimisation du risque structurel est liée à la définition de la dimension de Vapnik-Chervonenkis, appelée *dimension VC* [Vap95], permettant de caractériser la « richesse » d'une famille de fonctions de séparation. Choisir un ensemble de fonctions trop large conduit à des risques de surapprentissage, tandis qu'une famille trop restreinte peut ne pas contenir de solution satisfaisante. Définissons le risque empirique comme étant l'erreur moyenne de classification, calculée sur la base d'apprentissage. L'approche de minimisation du risque structurel consiste à choisir, parmi les solutions minimisant le risque empirique, celle qui est dotée de la dimension VC optimale. On se ramène à un problème d'optimisation quadratique, dont l'interprétation géométrique, illustrée en figure 2.10, est simple : on recherche l'hyperplan maximisant la *marge*, c'est-à-dire la somme des distances aux plus proches exemples des deux classes. Ces plus proches observations sont appelées *vecteurs de support*.

On peut facilement étendre cette technique au cas non linéaire par l'utilisation d'une fonction de noyau, comme pour les techniques de projection statistique (voir section 2.2.2.6). Dans le cas non linéaire où aucun hyperplan ne peut séparer linéairement les deux classes, il faut *relâcher* les contraintes, c'est-à-dire permettre que certains des exemples soient du mauvais côté de la frontière déterminée par l'hyperplan. On parlera alors de *marge souple* [PCST00]. Dans le contexte de la reconnaissance de visages, le nombre de personnes à reconnaître est généralement supérieur à deux. Néanmoins, la généralisation des SVM au cas multi-classes est un problème complexe [PCST00]. Afin de contourner ces difficultés, la plupart des techniques sont basées sur une formulation binaire du problème.

Dans [Phi99], Phillips a utilisé la formulation des sous-espaces Bayésiens (*cf.* section 2.2.2.3) à la manière de Moghaddam et Pentland [MP97] pour réduire le problème à deux classes : les variations intra-classe et les variations extra-classe. Les résultats expérimentaux ont montré que l'utilisation des SVM pour la classification apporte une amélioration par rapport à une simple distance Euclidienne.

Jonsson *et al.* [JKLM00] proposent une solution pour l'authentification, consistant à construire un SVM spécifique pour chaque personne enregistrée, à partir de signatures obtenues par ACP ou par ADL. Ils montrent au travers de résultats expérimentaux que, pour classer des signatures

obtenues par ACP, les SVM sont plus performants qu'une simple distance Euclidienne ou de corrélation, mais que les SVM n'apportent pas d'amélioration dans le cas de l'ADL. Cela peut être expliqué par le fait que l'ADL suffit à séparer correctement les différentes classes dans l'espace de projection, la phase de maximisation de la marge devenant ainsi superflue.

2.3 Les approches locales ou hybrides

Dans cette section, nous présentons les principales approches *locales*, c.-à-d. basées sur l'étude de caractéristiques extraites localement de différentes régions des visages. Nous exposerons également des techniques *hybrides*, au sens où elles utilisent conjointement des caractéristiques globales et locales des visages.

En section 2.3.1, nous passerons en revue les approches basées sur l'extraction de caractéristiques géométriques, avant de présenter en section 2.3.2 la fusion de sous-espaces modulaires, appliquant localement des techniques de projection statistique globales vues en section précédente. Puis, nous étudierons les méthodes basées sur les Modèles de Markov Cachés en section 2.3.4, avant d'aborder en section 2.3.5 l'utilisation de graphes par le biais de méthodes telles que l'*Elastic Graph Matching*.

2.3.1 Les approches géométriques

Les approches géométriques font partie des plus anciennes techniques utilisées dans le cadre de la reconnaissance de visages. Elles consistent à extraire, entre autre paramètres, les positions relatives des caractéristiques faciales telles que les yeux, le nez, la bouche, etc. Par exemple, dans [BP93], Brunelli et Poggio utilisent un ensemble de trente-cinq éléments géométriques extraits automatiquement, illustrés en figure 2.11. Les caractéristiques extraites des visages sont comparées deux à deux à l'aide d'une distance de Mahalanobis.

Cette approche nécessite une très grande précision dans la détection des divers éléments faciaux, ce qui reste un problème difficile dans des conditions générales de prise de vue et une voie de recherche (voir section B). De plus, la plupart des caractéristiques extraites ne sont pas robustes à des changements d'expression faciale ou de pose de la tête. Dans [BP93], il est montré que les techniques modulaires, qui font l'objet de la section suivante, sont plus efficaces que cette approche géométrique.

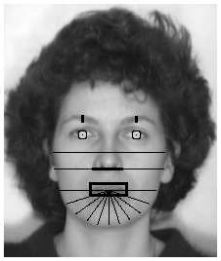


FIG. 2.11 – *Caractéristiques utilisées dans [BP93]. Celles-ci sont, entre autres : les dimensions du visage, les positions relatives des divers éléments faciaux et des segments de droite caractérisant la forme du menton.*

Les cartes de contour sont très utilisées dans le domaine de la reconnaissance de formes.

Elles présentent notamment l'avantage d'être robustes à des changements d'illumination dans les images. Elles ont été utilisées pour la première fois dans le contexte de la reconnaissance de visages par Takács dans [Tak98]. Cette approche consiste à comparer les images de visages par une mesure de similarité estimée directement entre leurs cartes de contour binaires, obtenues par le biais du filtre de Sobel. La mesure de similarité utilisée est inspirée de la distance de Hausdorff [HKR93], qui permet de comparer deux images sans pour autant nécessiter de mise en correspondance explicite des points issus de ces images.

Dans [GL02], cette approche a été améliorée par l'utilisation des lignes de contour des visages (au lieu de simples cartes de contour). Les lignes de contour sont obtenues en groupant les pixels de la carte de contour de manière à obtenir des segments de droite. Chaque visage est donc représenté par une carte appelée *Line Edge Map* (voir figure 2.12). On peut rapprocher cette

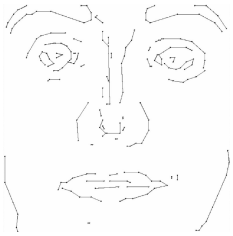


FIG. 2.12 – *Exemple de carte des lignes de contour d'un visage. Il s'agit de la signature utilisée pour la classification dans le contexte de la technique de Line Edge Map [GL02].*

technique des études biologiques qui ont montré la capacité du cerveau humain à reconnaître un visage depuis son dessin ou sa caricature [Per75]. La représentation d'un visage par LEM est moins coûteuse en termes de stockage que l'image initiale, puisque seules les positions des extrémités des segments de droite sont enregistrées. La distance utilisée pour la classification des visages est la même que dans [Tak98]. Les expérimentations reportées sur des images frontales montrent que la technique de LEM est plus efficace que celle de [Tak98], reposant sur des simples cartes de contour. Elle donne également des résultats de classification significativement meilleurs que les *eigenfaces* en présence de changements d'illumination. En revanche, elle est moins robuste aux variations d'expression faciale et d'angle de prise de vue, car ces deux facteurs jouent un rôle très important dans les signatures extraites.

2.3.2 Les techniques modulaires

Les méthodes détaillées dans cette section reposent sur des approches globales (présentées en section 2.2), appliquées de manière modulaire à différentes régions faciales, et combinées de manière à obtenir un modèle global alliant plusieurs modèles locaux. L'idée de ces approches est que les différentes régions faciales ne sont pas affectées de la même manière par les différentes sources de variabilité. Par exemple, le port de lunettes de soleil change considérablement l'aspect des yeux, tandis qu'un sourire affectera plus la région de la bouche. En considérant indépendamment des caractéristiques locales extraites de ces différentes régions faciales, on espère apporter une certaine robustesse, essentiellement vis-à-vis des changements d'expression faciale et des occultations partielles.

Brunelli et Poggio ont généralisé l'approche de la corrélation de Baron [Bar81] (voir section 2.2.1) à plusieurs régions faciales [BP93]. Quatre régions sont considérées : les yeux, le nez, la bouche et la région faciale dans sa globalité (du haut des sourcils jusqu'au menton). Lorsqu'un visage-requête doit être reconnu, on commence par le segmenter en régions à la manière de la

base d'apprentissage, puis on applique la technique basée sur la corrélation pour chaque région faciale. Les résultats sont combinés à l'aide d'un réseau HyperBF. Les résultats expérimentaux montrent que les caractéristiques faciales les plus discriminantes sont, en ordre décroissant de pouvoir discriminant : les yeux, le nez, la bouche et la globalité du visage.

Dans [PMS94], Pentland *et al.* ont introduit l'approche des *Modular Eigenspaces*. Les régions faciales retenues englobent la totalité du visage, les yeux et le nez. Une ACP est appliquée sur chacune de ces régions faciales et les résultats de classification obtenus sur chacune des régions sont agrégés. La bouche étant trop sensible à des changements d'expression faciale, sa prise en compte engendrerait une baisse des taux de reconnaissance. Cette approche peut être qualifiée d'hybride, puisqu'elle utilise à la fois des caractéristiques globales et locales. Pentland *et al.* ont montré qu'elle est plus efficace que les techniques globale et strictement locale (ne prenant pas en compte la globalité du visage) prises séparément.

En 2003, Heisele *et al.* ont introduit une technique modulaire utilisant les Machines à Vecteurs de Support (SVM). Dix caractéristiques faciales sont détectées, et on extrait les blocs de pixels englobant ces régions faciales. Chacun de ces blocs est transformé en un vecteur par concaténation des lignes (ou colonnes) de pixels. Les vecteurs correspondant aux dix caractéristiques sont concaténés pour obtenir un vecteur (de grande taille) par observation, ces vecteurs constituant les signaux d'entrée.

Dans [PG05], Price et Gee ont introduit une technique modulaire basée sur une variante de l'ADL alliant l'ADL Directe (méthode qui sera détaillée en section 3.3.3.3) et l'ADL Pondérée (qui sera détaillée en section 3.4.3.2). Les régions faciales considérées sont : la région faciale dans son ensemble, une bande faciale (de même largeur que la région faciale) s'étalant du front jusqu'au-dessous du nez, et une bande faciale contenant les yeux. Les résultats expérimentaux montrent que cette approche est plus performante que les techniques des *eigenfaces* et des *fisherfaces*, et notamment plus robuste aux changements dans les conditions d'illumination du visage, d'expression faciale et d'occultations partielles.

2.3.3 L'Analyse des Caractéristiques Locales

L'Analyse des Caractéristiques Locales (ACL) est une technique basée sur l'extraction par ACP de caractéristiques locales [PA96]. Contrairement aux *eigenfaces*, l'ACL est *topographique*, à savoir que les pixels voisins sont liés entre eux par des relations. Pour cela, l'ACL utilise des noyaux à support local, que l'on peut voir comme un ensemble de filtres locaux K_{G_i}, où les G_i sont les grilles (de support locaux). Ces grilles, ainsi que leurs noyaux associés, sont illustrés en figure 2.13. Une ACP est appliquée sur les images de la base d'apprentissage filtrées localement. On peut remarquer que le mode de mise en œuvre de l'ACP est proche de celui utilisé par Bartlett pour l'ACI dans le cadre de l'algorithme d'Architecture I [BMS02] (*cf.* section 2.2.2.5). La recherche du meilleur ensemble de grilles locales $G^* = \{G_1, \ldots, G_g\}$ se fait par la minimisation de l'erreur de reconstruction des images initiales. Selon les conclusions de Penev et Attick [PA96], la qualité perçue de la reconstruction est meilleure avec une ACL qu'avec une ACP. De plus, l'utilisation conjointe des *eigenfaces* et de l'ACL permet de diminuer l'erreur de reconstruction. L'utilisation conjointe de l'ACP et de l'ACL ouvre donc la voie à une technique hybride efficace de représentation des visages.

Cette approche souffre néanmoins de deux inconvénients majeurs : premièrement, la taille des signatures des visages est beaucoup plus importante qu'avec la méthode des *eigenfaces* et secondement elle repose sur une procédure d'optimisation itérative coûteuse et potentiellement instable. De plus, les caractéristiques sont choisies de manière à être le plus représentatives possible des images de visages, mais non dans un but de séparation des classes. Très peu d'informations ont

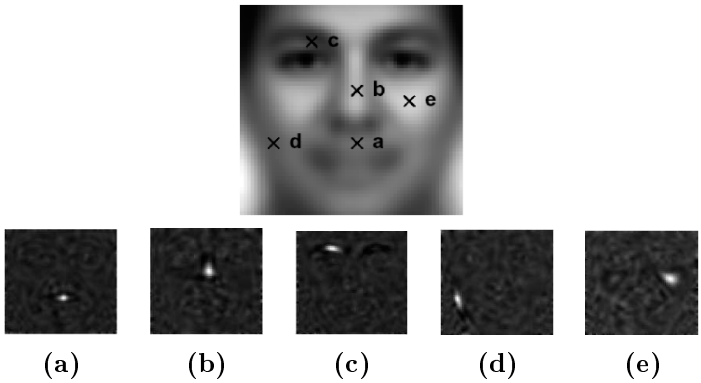


FIG. 2.13 – *Extrait de [PA96]. L'image en première ligne donne les positions sur le visage des centres des différentes grilles locales. Les images (a) à (e) de la seconde ligne montrent les noyaux locaux associés à ces grilles.*

été fournies concernant la manière dont ces caractéristiques pourraient combinées dans un but de classification. D'ailleurs, les articles de Penev et Atick [PA96] et de Penev [Pen00] ne reportent pas de résultats d'évaluation pour une tâche de reconnaissance. Néanmoins, le groupe Visionic revendique l'utilisation de l'ACL dans le logiciel commercial de reconnaissance de visages FaceIt (voir table 1.1).

2.3.4 Les Modèles de Markov Cachés

Les Modèles de Markov Cachés (MMC), initialement introduits par Baum *et al.* dans les années 1960, permettent de caractériser les propriétés statistiques d'un signal. D'abord utilisés en reconnaissance de la parole à partir des années 1980, ce n'est qu'en 1994 qu'ils furent introduits dans le cadre de la reconnaissance de visages par Samaria [Sam94].

Un MMC définit des variables aléatoires, formant une chaîne de Markov[9] cachée (c'est-à-dire non observable), dont la valeur dépend donc des éléments précédents dans une séquence. Ils forment une structure composée d'un nombre fini d'*états*, chacun étant associé à une densité de probabilité, d'une matrice A de *probabilités de transition* d'un état à l'autre, et d'une *distribution d'état initial* Π. Afin de caractériser précisément la séquence, on considère également un *alphabet de symboles*, associé à la matrice de probabilité B de ces symboles. Un MMC peut donc être caractérisé par le triplet (Π, A, B). Les MMC permettent ainsi de modéliser des séquences de symboles de manière dynamique (puisque tout élément influe sur la valeur des éléments qui le suivent dans la séquence).

Levin et Pieraccini [LP92] ont montré qu'une adaptation bidimensionnelle entièrement connectée des travaux réalisés dans le cadre du traitement unidimensionnel de la parole serait beaucoup trop complexe en temps de calcul. C'est pourquoi Samaria [Sam94], ainsi que Nefian [Nef99] ont commencé par introduire une structure simple de MMC unidimensionnel (MMC 1D), avant de proposer des algorithmes plus complexes, appelés pseudo bidimensionnels (MMC pseudo-2D), au sens où ils ne sont pas entièrement connectés dans les deux directions.

Pour pouvoir appliquer les MMC à la reconnaissance de visages, il faut définir un alphabet de symboles représentatif des visages. La première solution proposée par Samaria *et al.* [Sam94] (MMC 1D) consiste à segmenter les images de visages en un ensemble de régions (bandes faciales)

9. vérifiant les propriétés d'horizon limité et de stationnarité.

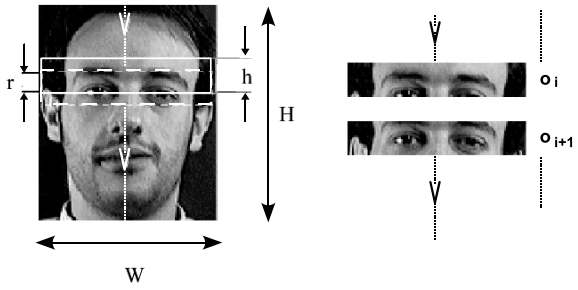


FIG. 2.14 – *Extrait de [Nef96]. L'image de visage, initialement de taille $H \times W$, est segmentée en bandes faciales de hauteur h pixels, avec un recouvrement r entre bandes. L'ordre d'apparition des observations o_i est de haut en bas.*

couvrant toute la largeur de l'image (voir figure 2.14). À chacune de ces bandes faciales est associé un symbole. Il existe cinq états, correspondant au front, aux yeux, au nez, à la bouche et au menton. Sous l'hypothèse que le visage est représenté dans une position frontale, un ordre d'apparition naturel est *de haut en bas* (par analogie avec le modèle *de gauche à droite* retenu dans le cadre du traitement de la parole). Les observations émises par chacun de ces états sont : soit les vecteurs obtenus par concaténation des lignes de pixels des bandes faciales, soit les vecteurs de leur compression via une transformation en cosinus discrète bidimensionnelle (DCT) [RY90]. Un modèle MMC est construit pour chacune des personnes (classes) de la base d'apprentissage, en utilisant classiquement un algorithme de Viterbi [For73] pour l'initialisation des paramètres et la méthode de Baum-Welch [Bau72] pour leurs réajustements. Lorsqu'un visage-requête est présenté au système, on segmente l'image à la manière de la base d'apprentissage, puis on calcule les vraisemblances de chacun des modèles de Markov (avec l'algorithme de Viterbi) pour finalement lui assigner l'identité associée au modèle le plus vraisemblable. Cette technique a deux inconvénients majeurs. Tout d'abord, elle nécessite une très bonne précision lors de la segmentation du visage en sous-bandes, ce qui est une tâche difficile. De plus, le sens de parcours (du haut vers le bas) permet de modéliser des déformations verticales des visages, mais n'est pas robuste aux variations horizontales telles que les rotations de la tête en profondeur (vers la gauche ou la droite).

C'est pourquoi Samaria [Sam94] et Nefian [Nef99] utilisent le modèle de MMC pseudo 2D, aussi appelé *MMC planaire*. Cette technique repose sur la définition de *super-états*, eux-mêmes Markoviens, décrivant l'ensemble des états de Markov 1D. Les bandes faciales évoquées précédemment sont découpées en sous-régions (blocs de pixels) décrites par des MMC 1D, avec une transition de gauche à droite, et définissent les *super-états*, d'ordre d'apparition de haut en bas. Par conséquent, le réseau n'est pas entièrement connecté dans les deux directions. Plus récemment, Nefian a introduit une technique généralisant les MMC pseudo 2D, remplaçant les MMC par un réseau Bayésien plus général, et a prouvé l'efficacité de cette méthode pour la reconnaissance de visages [Nef02].

Perronnin [Per04] a introduit un modèle capable d'injecter dans le cadre probabiliste des MMC 2D (si les états considérés sont discrets) et des Modèles Espace-État (MME) 2D (dans le contexte d'états continus), un ensemble de transformations locales conçues de telle manière que les déformations voisines restent cohérentes entre elles. L'utilisation de ces transformations locales

vise à estimer l'ensemble des déformations globales des visages, supposé trop complexe pour être modélisé directement. Des algorithmes performants d'approximation, appelés *turbo MMC* et *turbo MME*, sont proposés pour la mise en œuvre de ces méthodes. Les résultats expérimentaux montrent que, dans le cadre du problème de l'authentification, la seconde approche est très performante, et supérieure à la première. La technique proposée est conçue pour être robuste aux changements dans la pose, l'expression faciale et les conditions d'illumination.

2.3.5 Les approches basées sur les graphes

Les techniques présentées dans cette partie sont basées sur la mise en correspondance de graphes. Pour un introduction détaillée de ce problème, se référer à [Jol01]. Nous nous intéressons particulièrement aux approches appelées *Elastic Graph Matching* et *Elastic Bunch Graph Matching*. Ces techniques sont basées sur la méthodologie d'*Architecture de Lien Dynamique*. Cette dernière est étroitement liée à la théorie des réseaux de neurones ; basée sur l'élasticité des liaisons synaptiques, elle permet de structurer par le biais de graphes des neurones caractérisant des propriétés locales des visages.

2.3.5.1 L'*Elastic Graph Matching*

La technique dite d'*Elastic Graph Matching* (EGM) a été initialement introduite par Lades *et al.* en 1993 [LVB+93]. À chaque image de la base d'apprentissage est associé un graphe qui lui est propre. On utilise pour cela une grille régulière, placée sur les images de visages. Les caractéristiques extraites sont généralement des coefficients de Gabor ou des vecteurs de propriétés morphologiques [KTP00]. Le treillis de la grille utilisée pour les images-requêtes est généralement plus fin que pour les images d'apprentissage. La distance entre l'image-requête et une image connue est définie comme étant la meilleure mise en correspondance $\mathcal{M}^*$ entre les vecteurs de caractéristiques des deux images (de tailles différentes), parmi les solutions possibles $\mathcal{M}$. On restreint le champ de recherche aux solutions préservant un certain nombre de contraintes spatiales. On définit pour cela la fonction de coût suivante :

$$C(\mathcal{M}) \quad = \quad C_l(\mathcal{M}) + \rho C_g(\mathcal{M}) \tag{2.17}$$

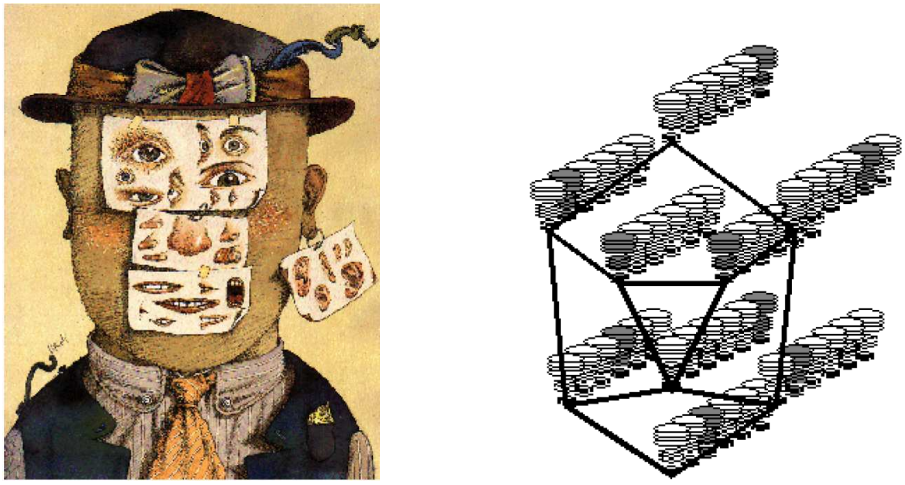
où $C_l(\mathcal{M})$ est la somme des coûts locaux de la mise en correspondance $\mathcal{M}$ des caractéristiques deux à deux, et $C_g(\mathcal{M})$ est le coût de déformation global du modèle. Le paramètre ρ est une mesure de la *rigidité* du graphe ; sa valeur est généralement fixée *ad hoc*. La distance entre deux vecteurs de caractéristiques est mesurée via la distance du cosinus (*cf.* Annexe D).

Étant donné la combinatoire du modèle, le nombre de solutions possibles est très important, et ceci même pour des tailles de treillis modérées. Il est par conséquent impossible de mener une recherche exhaustive. C'est pourquoi un algorithme en deux étapes a été proposé. Après avoir initialisé $\mathcal{M}$ à $\mathcal{M}_0$, où $\mathcal{M}_0$ correspond à un graphe totalement rigide ($\rho \to \infty$), les positions des nœuds du graphe sont localement perturbées jusqu'à ce que l'on atteigne une solution $\mathcal{M}^*$ minimisant localement la fonction de coût.

Des améliorations ont été apportées plus tard à ce modèle [DJK+02]. Dans [KTP00] Kotropoulos *et al.* cherchent à réduire l'influence de l'initialisation de l'algorithme (pouvant engendrer la convergence vers des minima locaux) par une procédure probabiliste fournissant, à chaque itération, le couple optimal de transformations globales et locales. Dans [TKP01], les auteurs choisissent de neutraliser le terme C_g dans l'équation (2.17), et d'utiliser alternativement un ensemble de contraintes locales pour éviter des déformations improbables du visage. Les hyperplans de séparation optimaux sont déterminés, pour chaque caractéristique, par l'utilisation de SVM.

2.3.5.2 L'*Elastic Bunch Graph Matching*

La technique d'*Elastic Bunch Graph Matching* (EBGM) [WFKvdM97] est proche de celle de l'*Elastic Graph Matching*. La différence avec l'EGM est que l'EBGM utilise un même graphe pour la modélisation de tous les visages, ce qui semble cohérent du fait de la structure géométrique prédéfinie des images de visages. Chaque nœud est associé à une caractéristique faciale (yeux, nez, bouche) ou à des points de contour. Au lieu de construire un modèle pour chaque image, on construit donc un modèle général de représentation, appelé *Face Bunch Graph* (FBG), depuis l'intégralité de la base d'apprentissage (voir figure 2.15). Tous les vecteurs correspondant à un même nœud sont regroupés de manière à représenter l'ensemble des états possibles de ce nœud. Le but est d'incorporer dans chaque nœud le plus de variabilité possible, en utilisant notamment des images différant dans l'expression faciale. Les nœuds sont représentés indépendamment les uns des autres, ce qui confère à l'EBGM un pouvoir combinatoire important et une bonne capacité de généralisation (p. ex. connaissant deux vues : l'une où les deux yeux sont ouverts et l'autre où les yeux sont fermés, on devrait être capable de reconnaître la personne si elle cligne d'un œil).



(a) **(b)**

FIG. 2.15 – *Un Face Bunch Graph (FBG) représenté d'un point de vue (a) artistique et (b) scientifique. Le FBG vise à représenter l'ensemble des états possibles, pour chaque nœud.*

La base d'apprentissage est annotée à la main de manière à connaître les positions exactes des caractéristiques faciales et à construire un modèle qui soit le plus précis possible. Lorsqu'un visage-requête doit être reconnu, l'algorithme d'EBGM est utilisé pour en localiser les caractéristiques faciales et construire son graphe associé. La fonction de coût utilisée, proche de celle de l'EGM (voir équation (2.17)), inclut en outre de l'information de phase, de manière à lever l'ambiguïté entre des caractéristiques dont les valeurs sont proches et à estimer les translations locales. Tout comme pour l'EGM, l'algorithme est en deux étapes : la première permet de compenser les distorsions globales du visage et la seconde les dissimilarités locales. Puisque les nœuds des graphes correspondent aux mêmes caractéristiques faciales, la mise en correspondance est simplifiée et repose sur la distance du cosinus entre caractéristiques.

Puisque le pouvoir discriminant des différentes caractéristiques faciales n'est pas le même (par exemple, on considère généralement que la bouche, plus sujette à des distorsions du fait des changements d'expression faciale, est une caractéristique moins discriminante que les yeux) et varie en fonction des sujets, certaines approches visent à pondérer l'importance de ces caracté-

ristiques pour la mise en correspondance de graphes [Krü97]. L'algorithme proposé dans [Krü97] est conçu pour être également robuste aux changements de pose. L'approche d'EBGM est performante dans le cadre de la reconnaissance de visages. Elle est néanmoins très coûteuse en termes de temps de calcul, tant pour la construction du modèle que pour la mise en correspondance de graphes.

2.4 Comparaison des performances des méthodes

2.4.1 Présentation des résultats expérimentaux

Les tables 2.1 et 2.2 fournissent une comparaison des taux de reconnaissance de la plupart des méthodes présentées dans ce chapitre. Les résultats de la table 2.1 sont obtenus sur les bases FERET et ORL, tandis pour la table 2.2 les bases utilisées sont les bases de Yale et AR. Ces bases de visages sont décrites en annexe A (p. 169).

Décrivons les principales évaluations desquelles sont tirés ces résultats. Les résultats faisant référence à [PMRR00] sont issus de l'évaluation FERET de Mars 1997 (*cf.* section 1.7). Dix systèmes ont été évalués, parmi lesquels deux algorithmes basés sur les *eigenfaces*, un système utilisant les sous-espaces Bayésiens, trois algorithmes reposant sur la mise en œuvre des *fisherfaces* et une technique basée sur l'Elastic Bunch Graph Matching (EBGM). La base de visages utilisée pour l'évaluation est un sous-ensemble de FERET. Les deux systèmes les plus performants sont l'EBGM et les *fisherfaces* de Zhao *et al.* [Zha99]. L'EBGM fournit de meilleurs résultats que les *fisherfaces* sur les vues duplicate et la vue fc ; pour la vue fb les *fisherfaces* sont légèrement meilleurs que l'EBGM. Il faut cependant noter que le protocole expérimental mis en œuvre n'est pas favorable aux *fisherfaces*, puisque seulement deux vues par personne sont utilisées pour l'apprentissage, ce qui est en général insuffisant pour un apprentissage efficace de l'ADL [MK01].

Moghaddam fournit dans [Mog02] une comparaison des techniques des *eigenfaces*, de l'Analyse en Composantes Indépendantes (ACI) et de la technique des sous-espaces Bayésiens. Ils utilisent cinq partitions aléatoires de la base FERET en une base d'apprentissage et une base de test contenant chacune 140 personnes. Les résultats expérimentaux montrent que les *eigenfaces* et l'ACI sont à peu près aussi performantes, mais que l'ACI est beaucoup plus instable que l'ACP, au sens où ses performances varient beaucoup plus en fonction de la base considérée. Les sous-espaces Bayésiens sont significativement plus efficaces que les deux autres techniques.

Dans [GSC01], Gross *et al.* procèdent à l'évaluation sur plusieurs bases, dont la base AR, des techniques des *eigenfaces*, des *fisherfaces* et du logiciel commercial FaceIt. Ce dernier reposerait sur l'Analyse des Caractéristiques Locales (ACL) détaillée en section 2.3.3, mais nous ne disposons pas de beaucoup d'informations sur le mode de mise en œuvre et les prétraitements appliqués sur l'image. Plusieurs bases sont utilisées, et l'impact de différents facteurs est étudié, comme décrit en section 1.5. Pour l'évaluation, les algorithmes des *eigenfaces* et des *fisherfaces* sont entraînés sur une partie des bases utilisées, tandis que l'apprentissage de FaceIt a eu lieu en amont, sur une base à propos de laquelle nous ne disposons d'aucune information. Les résultats expérimentaux montrent que FaceIt est dans la majorité des cas supérieur à la méthode des *fisherfaces*, elle-même plus performante que la technique des *eigenfaces*. Par contre, dans le cas où les yeux sont occultés par des lunettes noires (base AR), les *fisherfaces* fournissent un taux de reconnaissance de 45% contre seulement 10% pour FaceIt. Si l'on ajoute à cette occultation des différences dans les conditions d'illumination, les taux de reconnaissance passent à 27% pour l'ADL et 6% pour FaceIt. La méthode mise en œuvre dans FaceIt est en revanche beaucoup plus tolérante à une occultation de la région basse des visages que les techniques des *eigenfaces* et

des *fisherfaces*. Le comportement des méthodes des *eigenfaces* et des *fisherfaces*, qui sont basées sur l'étude de l'ensemble des valeurs de pixels (méthodes globales) modélisés sous la forme de vecteurs-images (*cf.* section 2.2.2) est facilement interprétable. En effet, l'occultation d'une importante région du visage entraîne la modification des valeurs d'une grande partie de ces pixels et ne peut que résulter en une diminution importante des performances. *A contrario*, la technique locale d'Analyse des Caractéristiques Locales utilisée par le logiciel FaceIt doit sans aucun doute accorder une forte importance à la région supérieure du visage, si bien que l'algorithme est à la fois relativement tolérant à une occultation de la partie basse du visage, mais non robuste à des modifications de l'apparence des yeux. Néanmoins, il faut noter que le contenu des bases d'apprentissage influe beaucoup sur les performances des techniques statistiques, et que l'apprentissage de FaceIt n'a pas été effectué avec les mêmes bases que les deux autres méthodes, ce qui biaise la comparaison des taux de reconnaissance.

D'une manière générale, on peut remarquer que les bases utilisées pour l'évaluation ont un impact important sur les résultats obtenus. Ainsi, pour les *eigenfaces* et sur la base FERET, les taux de reconnaissance passent de 80% à 20% selon que l'on considère comme base de test l'ensemble fb ou fc (*cf.* annexe A).

Notons également le protocole d'évaluation retenu influe fortement sur les performances des systèmes. Par exemple, dans [Yan02], les performances des systèmes sont évalués selon une stratégie *leave-one-out*, détaillée ci-après. On dispose initialement d'une base de visages. On tire au hasard une image dans cette base. La base d'apprentissage est constituée de toutes les images de la base initiale, à l'exception de cette image tirée au hasard. L'algorithme est entraîné sur la base d'apprentissage, et l'on recueille son résultat de classification sur l'exemple isolé. Cette opération est répétée autant de fois qu'il y a d'images dans la base d'apprentissage (soit 400 fois sur la base ORL, décrite en annexe A). Les taux de reconnaissance extraits de [Yan02] et figurant dans le tableau 2.1 (base ORL) correspondent au ratio moyen d'exemples correctement classés dans ces conditions. L'apprentissage de chaque classifieur est donc mené depuis 9 à 10 images par personne. Puisque la base ORL ne contient pas de changement d'apparence très importants des visages, plus le nombre de vues par personne utilisées pour l'apprentissage est élevé, plus les techniques statistiques sont performantes. La technique d'évaluation *leave-one-out* permet donc de simuler des conditions très favorables aux techniques statistiques. Par conséquent, les performances reportées dans [Yan02] sur la base ORL sont surévaluées par rapport à celles fournies dans [LGTB97, JKLM00, Sam94, LKL97, EWLT02], où le nombre de vues d'apprentissage par personne est seulement de 5.

2.4.2 Conclusion

On retient des tableaux 2.1 et 2.2 les très bonnes performances des techniques statistiques de réduction de dimension et surtout des *fisherfaces* (mises en œuvre avec ou sans fonction de noyau) en comparaison avec les autres techniques de l'état de l'art. On remarque également qu'un mode de mise en œuvre modulaire de ces techniques statistiques semble apporter une amélioration supplémentaire. Notons de plus que le fait de remplacer la distance au plus proche voisin par un Réseau de Fonctions à Base Radiale pour la classification issue des *fisherfaces* apporte, sur la base ORL, un gain important dans les taux de reconnaissance [EWLT02].

2.5 Conclusion

Dans ce chapitre, nous avons passé en revue les principales techniques de reconnaissance automatique de visages proposées à ce jour. Ces méthodes se décomposent en deux grandes familles : les approches globales, pour lesquelles les caractéristiques sont extraites directement depuis l'ensemble des valeurs de pixels des images et les approches locales, basées sur l'extraction de signatures extraites localement du visage. Les techniques dites hybrides utilisent conjointement ces deux types de modélisation. Les approches hybrides, proches du fonctionnement du système visuel humain, sont généralement très performantes et plus robustes à des changements d'apparence du visage dus par exemple à des variations dans l'expression faciale que les techniques globales. Néanmoins, elles sont généralement plus coûteuses en temps de calcul, pour la phase d'apprentissage comme pour la classification.

Parmi les techniques globales, on compte notamment les méthodes basées sur la projection statistique, aussi appelées techniques de réduction de dimension. Elles visent à définir un espace de projection dans lequel les données sont projetées puis classées. Deux types de critères peuvent être utilisés pour déterminer ce sous-espace : un critère de représentativité des données (on cherche à préserver la distribution des données) ou de séparabilité en fonction de la classe d'appartenance. Pour optimiser le premier critère, on utilise essentiellement l'ACP (technique des *eigenfaces*) ou l'ACI tandis que, pour le second critère, l'ADL (méthode des *fisherfaces*) est généralement préférée. Ces techniques sont caractérisées par un apprentissage rapide, un faible nombre de paramètres à ajuster et de très bonnes performances. Nous avons notamment vu au travers de résultats expérimentaux provenant de différentes sources que la méthode des *fisherfaces* est très performante, en comparaison avec les autres techniques de l'état de l'art. Les *fisherfaces* présentent de plus l'avantage d'être facilement généralisables à des problèmes plus complexes : on peut par exemple les étendre à des problèmes non linéaires par l'utilisation d'une fonction de noyau. Ces très bonnes propriétés peuvent expliquer en partie le très fort intérêt relevé ces dernières années pour les techniques basées sur l'ADL dans le contexte de la reconnaissance automatique de visages. La plupart des solutions proposées dans ce cadre seront passées en revue dans le chapitre suivant.

Base	Source	Méthode	Taux de reconnaissance	Commentaires
FERET	[PMRR00]	corrélation [BM95]	83% et 5%	Base d'apprentissage : 3323 images de 1196 personnes. Bases de test : fb et fc (*cf.* annexe A)
		eigenfaces [TP91]	80% et 20%	
		fisherfaces [Zha99]	96% et 59%	
		EBGM [WFKvdM97]	95% et 81%	
	[Mog02]	*eigenfaces* [TP91]	77%	5 bases d'apprentissage : chacune incluant 140 images de 140 personnes. 5 bases de test : chacune avec 140 images ou +
		ACI [BMS02]	77%	
		ACP à noyau [Yan02]	87%	
		ACP Bayésienne [Mog02]	95%	
	[PMS94]	sous-espaces modulaires [PMS94]	95%	7562 images de 3000 pers.
ORL	[Yan02]	*eigenfaces* [TP91]	97,5%	Stratégie d'évaluation *leave-one-out*
		fisherfaces [ZCK98]	98,5%	
		ACI [BMS02]	93,75%	
		SVM	97%	
		eigenfaces à noyau [Yan02]	98%	
		fisherfaces à noyau [Yan02]	98,75%	
	[LGTB97]	Kohonen+rdn convolutionnel [LGTB97]	96,2%	partitions aléatoires en une base d'apprentissage avec 5 images/pers. et une base de test avec 5 images/pers.
	[JKLM00]	SVM [JKLM00]	91,21%	
	[Sam94]	MMC1D [Sam94]	87%	
		MMC pseudo2D [Sam94]	95%	
	[LKL97]	rdn probabiliste [LKL97]	96%	
	[HLLM02]	*fisherfaces* [BHK97]	94,19%	
	[EWLT02]	*fisherfaces*+ FBR [EWLT02]	98,1%	

TAB. 2.1 – *Comparaison des performances des algorithmes proposés, sur les bases de visages FERET et ORL décrites en annexe A.*

Base	**Source**	**Méthode**	**Taux de reconnaissance**	**Commentaires**
Yale	[Yan02]	*eigenfaces* [TP91]	71,5%	Stratégie d'évaluation *leave-one-out*
		fisherfaces [ZCK98]	91,5%	
		ACI [BMS02]	71,5%	
		SVM	82%	
		eigenfaces à noyau [Yan02]	75,8%	
		fisherfaces à noyau [Yan02]	93,9%	
AR	[Per04]	MMC2D [Per04]	89%	évalué en moyenne sur les 3 variations d'expressions de la base
	[GSC01]	*eigenfaces* [TP91]	70,7%	
		fisherfaces [BHK97]	81,7%	
		FAceIt (ACL)	88%	
	[GL02]	Line Edge Map [GL02]	96,4%	

TAB. 2.2 – *Comparaison des performances des algorithmes proposés, sur les bases visages internationales Yale et AR, décrites en annexe A.*

Chapitre 3

Analyse Discriminante Linéaire et reconnaissance automatique de visages

3.1 Introduction

Nous avons vu au chapitre 1. que la reconnaissance de visages est l'un des domaines les plus actifs de l'analyse d'images et de la reconnaissance de formes. De multiples techniques, passées en revue au chapitre 2., ont été proposées dans ce contexte. L'Analyse Discriminante Linéaire (ADL) est, en particulier, une technique très prisée des chercheurs. La figure 3.1 montre l'évolution dans le temps du nombre de publications référencées dans la base IEEE Xplore (voir note de bas de page n° 2 en p. 5) ayant trait à l'utilisation de l'ADL pour la reconnaissance de visages. On constate un regain d'intérêt pour ces techniques au cours des dernières années et notamment

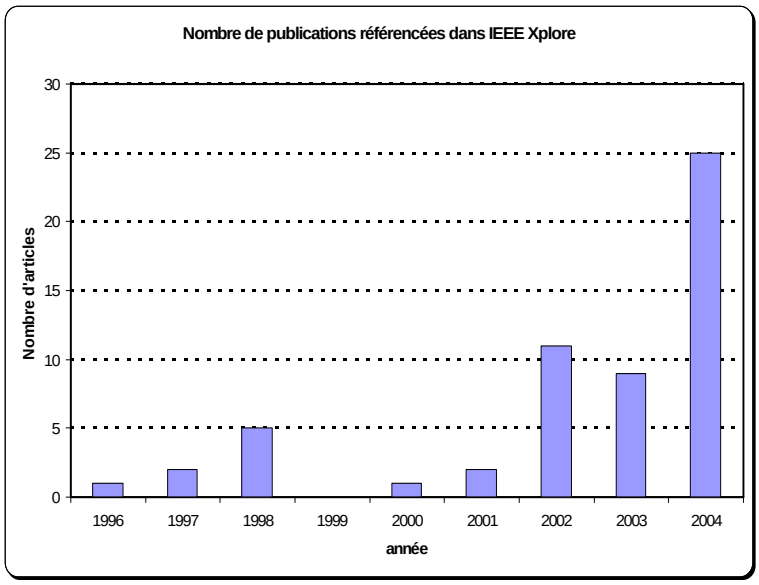


FIG. 3.1 – *Nombre de publications référencées par le moteur de recherche IEEE Xplore, en fonction de l'année de publication. Les titres des articles comptabilisés contiennent au moins l'un des termes-clés suivants : Face Recognition, Face Identification, Face Authentication et Face Verification, ainsi que Fisher, LDA, FLD ou Discriminant. On peut noter une augmentation importante du nombre de publications ces dernières années.*

en 2004. Malgré la multitude des familles de solutions proposées (voir chapitre 2.), on remarque que les techniques basées sur l'ADL représentaient en 2004 un huitième du nombre total de publications sur la reconnaissance de visages (voir figure 1.1).

Comme nous l'avons vu en section 2.2.2.4, lorsque l'on cherche à appliquer directement une ADL sur les images de visages (représentées de manière 1D par le biais de vecteurs), on est confronté au problème de la singularité. Celui-ci est dû à une sous-représentation des données en entrée et se retrouve dans la plupart des applications de reconnaissance de formes. Une part importante des efforts de recherche a porté sur la conception de solutions à ce problème. Il faut également noter que l'ADL repose sur un ensemble d'hypothèses dont la non-validité peuvent affecter les performances du processus. Une autre voie de recherche repose sur la définition d'algorithmes plus robustes à la violation de ces conditions. Enfin, dans le cas où les distributions des données sont trop complexes pour qu'une combinaison linéaire des caractéristiques permette de séparer correctement les classes, des variantes non linéaires de l'ADL, basées sur l'utilisation d'une fonction de noyau, ont été introduites.

Ce chapitre vise à passer en revue les principales méthodes issues de l'Analyse Discriminante Linéaire et appliquées à la reconnaissance de visages. En section 3.2 nous décrirons la technique d'Analyse Discriminante Linéaire standard. Le problème de la singularité et les solutions proposées sont présentés en section 3.3. Les variantes de l'ADL visant à la rendre plus robuste à la non-validité des hypothèses sous-jacentes sont détaillées en section 3.4. Enfin, l'utilisation d'une fonction de noyau dans le contexte de l'ADL sera étudié en section 3.5.

3.2 L'Analyse Discriminante Linéaire

3.2.1 Introduction

Les méthodes de discrimination, initialement introduites par Fisher [Fis36] et par Mahalanobis [Mah36] ont pour but de décrire et de classer des individus (aussi appelés observations ou exemples) caractérisés par un nombre important de variables prédictives, la plupart du temps numériques. Dans le contexte de la reconnaissance de visages, il s'agit de prédire l'identité d'une personne à l'aide d'une ou plusieurs images de son visage. Les variables prédictives, appelées par la suite *caractéristiques*, peuvent par exemple être les valeurs des pixels (en niveau de gris).

Considérons une population d'individus connus, formant un nuage de points et appelée *base d'apprentissage*. Cette population est partitionnée en k groupes (*classes*) constituant des sous-nuages, chaque classe correspondant à une identité différente. Chacun des individus est décrit par n caractéristiques et l'on connaît sa classe d'appartenance. L'Analyse Discriminante Linéaire (ADL) est donc une technique *supervisée*. Le but de l'ADL est de construire, à partir de ces données, un sous-espace linéaire de l'espace initial des données, dans lequel les k groupes sont le mieux séparés possible. Pour déterminer l'identité d'un nouvel arrivant, il suffit de le projeter dans ce sous-espace et de déterminer dans quel sous-nuage il repose (ou de quel sous-nuage il est le plus proche).

L'Analyse Discriminante Linéaire peut être vue comme une méthode *linéaire*, en ce sens qu'elle suppose que les différents groupes soient linéairement séparables dans un sous-espace de l'espace initial des données. Elle permet d'estimer la combinaison linéaire des caractéristiques permettant de séparer au mieux les k catégories d'individus.

Dans le contexte de la reconnaissance de visages, l'objectif de l'ADL est la classification : le modèle doit être capable, à partir des caractéristiques d'un nouvel arrivant, d'assigner à cet

individu une classe d'appartenance. L'ADL est donc une technique d'analyse de données capable de construire un *classifieur* (aussi appelé modèle). Le classifieur ainsi obtenu est entièrement déterminé par le sous-espace discriminant (dont les vecteurs de projection sont appelés *facteurs discriminants*) et la connaissance des exemples de la base d'apprentissage étiquetés par leur classe d'appartenance.

Cette section est organisée comme suit. La section 3.2.2 détaille les notations utilisées dans la suite du chapitre 3. La section 3.2.3 décrit la technique d'Analyse Discriminante Linéaire. Enfin, en section 3.2.4, nous discuterons des avantages et des limitations de l'ADL dans le contexte de la reconnaissance automatique de visages.

3.2.2 Données et notations

Le tableau 3.1 résume les notations utilisées dans ce chapitre.

Notation	Description	Notation	Description
Ω	ensemble des exemples	Ω_j	classe j
n	dimension des exemples	k	nombre de classes
N	nombre d'exemples de Ω	N_j	nombre d'exemples de Ω_j
A	matrice des observations de Ω ($A \in \mathbb{R}^{n\times N}$)	$A^{(j)}$	matrice des observations de Ω_j ($A^{(j)} \in \mathbb{R}^{n\times N_j}$)
$\overline{A}$	moyenne des exemples de Ω	$\overline{A_j}$	moyenne des exemples de Ω_j
S_T	matrice de variance totale de Ω	V_j	matrice de variance de Ω_j
S_b	matrice de variance inter-classe	S_w	matrice de variance intra-classe
W	matrice des facteurs discriminants	I_n	matrice identité de taille $n \times n$
0_n	matrice nulle de $\mathbb{R}^{n\times n}$	$0_{n\times 1}$	vecteur nul de $\mathbb{R}^n$

TAB. 3.1 – *Principales notations du chapitre 3*

On dispose d'une base d'apprentissage Ω composée de N individus (exemples) A_l de $\mathbb{R}^n$, chacun étant affecté à l'une des k classes. Le nuage de points Ω est donc partagé en k sous-nuages $\Omega_1, \Omega_2, \ldots, \Omega_k$, de matrices de variance $V_1, V_2, \ldots, V_k$. Considérons que les N individus sont affectés des poids $p_1, p_2, \ldots, p_N$, le poids q_j de chaque classe Ω_j est alors :

$$q_j = \sum_{A_l \in \Omega_j} p_l, \quad j = 1, \ldots, k \tag{3.1}$$

Généralement, on pose $p_1 = p_2 = \cdots = p_N = \frac{1}{N}$ et les q_j représentent les proportions de chacune des classes Ω_j dans la population totale Ω. Le centre de gravité $\overline{A_j}$ et la matrice de variance V_j de chaque classe Ω_j sont :

$$\overline{A_j} = \frac{1}{q_j} \sum_{A_l \in \Omega_j} p_l A_l \quad \text{et} \quad V_j = \frac{1}{q_j} \sum_{A_l \in \Omega_j} (A_l - \overline{A_j})(A_l - \overline{A_j})^T \tag{3.2}$$

Appelons *matrice de variance intra-classe* et notons S_w la moyenne pondérée des matrices V_j :

$$S_w = \sum_{j=1}^{k} q_j V_j = \sum_{j=1}^{k} p_l \sum_{A_l \in \Omega_j} (A_l - \overline{A_j})(A_l - \overline{A_j})^T$$

La *matrice de variance inter-classe* S_b mesure la variance (pondérée) des k centres de gravité :

$$S_b = \sum_{j=1}^{k} q_j (\overline{A_j} - \overline{A})(\overline{A_j} - \overline{A})^T$$

où $\overline{A} = \sum_{j=1}^{k} q_j \overline{A_j}$ est le centre de gravité de Ω.
La *matrice de variance totale* S_T vérifie la relation suivante (théorème de Huygens) :

$$S_T = S_w + S_b \tag{3.3}$$

Dans le cas où chaque individu est affecté du même poids $p_l = \frac{1}{N}$, et en introduisant les effectifs $N_1, N_2, \ldots, N_k$ des classes $\Omega_1, \Omega_2, \ldots, \Omega_k$, on obtient :

$$S_w = \frac{1}{N} \sum_{j=1}^{k} \sum_{A_l \in \Omega_j} (A_l - \overline{A_j})(A_l - \overline{A_j})^T \tag{3.4}$$

$$S_b = \frac{1}{N} \sum_{j=1}^{k} N_j (\overline{A_j} - \overline{A})(\overline{A_j} - \overline{A})^T \tag{3.5}$$

$$S_T = \frac{1}{N} \sum_{l=1}^{N} (A_l - \overline{A})(A_l - \overline{A})^T \tag{3.6}$$

Notons A la matrice de $\mathbb{R}^{n \times N}$ contenant les observations centrées : $A = [A_1 - \overline{A}, \ldots, A_N - \overline{A}]$.
Dans la suite, nous supposerons que les individus sont centrés, c'est-à-dire que $\overline{A} = 0$

3.2.3 Description simplifiée de l'ADL

Cette section vise à décrire de manière simplifiée l'ADL. Plus de détails sont donnés en annexe E (p. 187).

On recherche les directions de projection W_i, appelées *facteurs discriminants* et en nombre $g \leq k - 1$, correspondant à des directions de $\mathbb{R}^n$ qui séparent le mieux possible en projection les k groupes d'observations (voir figure 3.2). Ces g vecteurs W_i, de longueur n, définissent le *sous-espace discriminant* noté $\mathcal{F}$. La matrice $W = [W_1, W_2, \ldots, W_g]$ contenant les facteurs les plus discriminants définit la matrice de projection sur le sous-espace $\mathcal{F}$. La projection A'_l de $A_l \in \mathbb{R}^n$ sur W est donnée par :

$$A'_l = W^T A_l \tag{3.7}$$

Le vecteur A'_l ainsi obtenu, de longueur g, définit la *signature* associée à l'exemple A_l par l'ADL. On considérera qu'un sous-espace $\mathcal{F}$ est discriminant s'il minimise par projection les variations à l'intérieur des classes, tout en maximisant les variations entre classes. Les quatre critères suivants

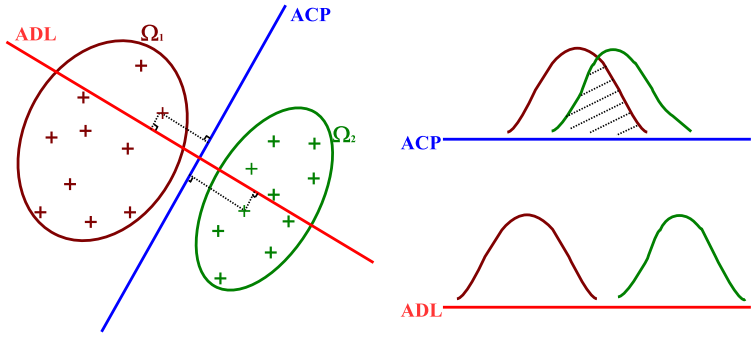


FIG. 3.2 – *Cas binaire : on recherche la droite de $\mathbb{R}^2$ telle que les groupes projetés sur cette droite soient le mieux séparés possible. L'axe ACP, qui est l'axe principal usuel, ne permet pas de séparer en projection les deux classes. Par contre, l'axe ADL possède un bon pouvoir discriminant.*

sont parmi les plus utilisés pour mesurer le pouvoir discriminant d'un sous-espace $\mathcal{F}$:

$$J_1(W) = \text{trace}(S_T'^{-1} S_b') \tag{3.8}$$

$$J_2(W) = \text{trace}(S_w'^{-1} S_b') \tag{3.9}$$

$$J_3(W) = \frac{|S_b'|}{|S_T'|} \tag{3.10}$$

$$J_4(W) = \frac{|S_b'|}{|S_w'|} \tag{3.11}$$

où $|X|$ est le déterminant de la matrice X et les matrices S_w', S_b' et S_T' sont respectivement les matrices de variance intra-classe, de variance inter-classe et de variance totale des vecteurs A_l' des données de la base d'apprentissage projetées sur $\mathcal{F}$. Il est facile de montrer que celles-ci peuvent s'écrire :

$$S_w' = W^T S_w W \qquad \text{et} \qquad S_b' = W^T S_b W \qquad \text{et} \qquad S_T' = S_w' + S_b' \tag{3.12}$$

où les matrices S_w et S_b sont respectivement les matrices de covariance intra-classe et inter-classe, estimées depuis la base d'apprentissage et données en équations (3.4) et (3.5). Il n'existe pas réellement d'indicateur permettant d'orienter son choix vers un critère plutôt qu'un autre. En revanche, le critère le plus couramment utilisé dans le domaine de la reconnaissance de formes est le critère (3.11), souvent appelé *critère de Fisher* [DHS01].

Nous montrons en section E.1.1 de l'annexe E que, sous l'hypothèse que la matrice S_w est inversible, les colonnes W_i de la matrice $W = [W_1, W_2, \ldots, W_g]$ maximisant le critère (3.11) sont les vecteurs propres de la matrice $S_w^{-1} S_b$ associés aux plus grandes valeurs propres non nulles. Nous montrons également que la matrice $S_w^{-1} S_b$ admet au plus $k-1$ vecteurs propres correspondant à des valeurs propres non nulles et que la valeur propre associée à chacun de ces vecteurs propres est une mesure pessimiste du pouvoir discriminant du vecteur propre associé. Traditionnellement, le nombre g de facteurs discriminants considérés augmente avec le nombre N d'observations dont on dispose [DHS01].

Étant donné que la matrice $S_w^{-1} S_b$ n'est pas nécessairement symétrique, le calcul de son système propre est potentiellement instable. Une technique basée sur les diagonalisations des matrices S_w et S_b, symétriques réelles, appelée *algorithme de Fukunaga*, est détaillée en section E.3.2 (p. 194) de l'annexe E. Cette méthode consiste à maximiser le déterminant de la

matrice $W^T S_b W$ sous la contrainte $W^T S_w W = I_g$, ce qui est équivalent à maximiser le critère (3.11) [Fuk90]. Résumons ici l'algorithme, qui comporte deux étapes. Dans une première phase on détermine le système propre (V, D_V) de la matrice S_w (où V est la matrice contenant en colonnes les vecteurs propres et D_V est la matrice diagonale contenant les valeurs propres associées), puis on projette les centres des classes $\overline{A_j}$ sur $VD_V^{-1/2}$, selon $\overline{A_j}'' = D_V^{-1/2} V^T \overline{A_j}$. La matrice de covariance intra-classe des données projetées est dite « blanchie », au sens où si l'on note S_w'' cette matrice et $Z = VD_V^{-1/2}$, celle-ci vérifie $Z^T S_w'' Z = I_n$. Dans un second temps on applique une ACP sur ces centres projetés $\overline{A_j}''$. On obtient ainsi la matrice B, contenant en colonnes les g facteurs principaux orthonormés et associés à des valeurs propres non nulles. La matrice de projection W de l'ADL est alors définie comme suit : $W = VD_V^{-1/2}B$. Remarquons que l'algorithme de Fukunaga nécessite que la matrice S_w soit régulière (car sinon certaines valeurs propres sont nulles et la matrice $D_V^{-1/2}$ n'est pas définie). Dans le cas contraire, l'ADL ne peut être appliquée.

Les données sont classées après projection dans $\mathcal{F}$, par une distance Euclidienne à la plus proche moyenne entre leurs signatures A_l' (*cf.* section E.1.2 de l'annexe E). Nous montrons en section E.2 de cette même annexe que, sous les hypothèses de multinormalité et d'*homoscédasticité* [10] des données en entrée, le classifieur construit par ADL est optimal au sens de la règle de Bayes. Si seule l'hypothèse de multinormalité est vérifiée, le classifieur Bayésien optimal est obtenu à l'aide d'une technique nommée Analyse Discriminante Quadratique (ADQ), qui repose sur l'estimation précise des matrices de variance V_j de chacune des classes et l'utilisation de règles quadratiques (*cf.* section E.2.2, p. 193 de l'annexe E).

3.2.4 L'ADL pour la reconnaissance de visages : avantages et limitations

L'Analyse Discriminante Linéaire est basée sur des fondements géométriques et probabilistes bien établis, que l'on peut aisément interpréter et, éventuellement, adapter au contexte (voir les nombreuses variantes dans la suite de ce chapitre). Puisque de plus l'Analyse Discriminante donne généralement de très bons résultats pour la tâche de reconnaissance de formes en général [McL04], et la reconnaissance de visages [BHK97, DY03, HLLM02, CNWB05, LPV03b, YY01, Yan02], les techniques basées sur l'ADL sont très prisées pour ce problème. Néanmoins, il existe un certain nombre d'écueils pour la mise en œuvre directe de l'ADL sur des vecteurs-images de visages, liés essentiellement à des problèmes de sous-représentation des données et de robustesse aux données aberrantes. Il faut également envisager le cas où les hypothèses sous-jacentes à l'ADL ne sont pas vérifiées.

Supposons que chaque image de visage soit constituée de $h \times w$ valeurs de pixels en niveaux de gris. Dans la plupart des approches globales de projection statistique (*cf.* section 2.2.2), ces matrices de pixels sont préalablement transformées en vecteurs par concaténation des lignes (ou des colonnes) de pixels. On parle de représentation unidimensionnelle (1D) des données. La base d'apprentissage est donc constituée de N vecteurs-images de taille $n = hw$ généralement très grande. À l'inverse, le nombre N d'exemples dont on dispose est la plupart du temps limité, et très faible en comparaison de leur dimensionnalité n ($N \ll n$). Construire un modèle d'ADL à partir de tels échantillons peut poser des problèmes de singularité et d'instabilité qui seront, ainsi que leurs solutions les plus usuelles, détaillés en section 3.3.

10. Hypothèse selon laquelle les matrices de variance des différentes classes sont égales.

Nous avons vu que, sous les conditions de multinormalité et d'homoscédasticité, l'ADL est optimale au sens de la règle de Bayes. S'il existe des outils pour tester les hypothèses de multinormalité et d'homoscédasticité [Sap90], ceux-ci sont d'autant plus difficile à mettre en œuvre et moins fiables que la taille des observations est importante (et que le nombre d'observations est petit relativement à cette taille). Ces hypothèses n'ont, à notre connaissance, jamais été vérifiées pour des bases de visages usuelles. Les performances de l'ADL sont plus ou moins affectées par des violations de ces conditions.

L'ADL est relativement robuste à des écarts à la normalité, pourvu que ceux-ci soient causés par des dissymétries et non par *des valeurs aberrantes* [TF96]. Par conséquent, les variantes non-normales de l'Analyse Discriminante sont très peu utilisées dans le cadre de la reconnaissance de visages. Certaines d'entre elles seront évoquées en section 3.4.2. Par contre, l'ADL est généralement très sensible à l'inclusion de données aberrantes. Il faudra donc apporter un soin particulier au choix des images à inclure dans la base d'apprentissage. Il est notamment d'usage de précisément « normaliser » les vues de visages, au sens où seule une région faciale bien délimitée est considérée (*cf.* section 1.6). Néanmoins, aucune précaution de ce type ne peut nous assurer que les données ne seront pas contaminées par de telles observations. Des variantes de l'ADL, conçues pour être plus résistantes aux observations aberrantes, sont présentées en section 3.4.3.

L'ADL est également sensible à la non-homoscédasticité des données en entrée. Si l'on soupçonne les données d'être hétéroscédastiques, faut-il pour autant mettre en œuvre une Analyse Discriminante Quadratique (voir section E.2.2 (p. 193) de l'annexe E), à la place de l'ADL? La réponse est en général non, car l'usage de règles quadratiques implique l'estimation de beaucoup plus de paramètres que la règle linéaire (il faut estimer précisément les k matrices de variance V_j) et que, bien souvent, le nombre d'observations n'est pas suffisant pour estimer précisément ces paramètres. Une mesure simple pour favoriser l'homoscédasticité consiste à sélectionner les vues à inclure dans la base d'apprentissage de manière à ce que chaque classe soit soumise au même type de variations. Par exemple, il est d'usage d'éviter de constituer une classe avec deux vues différant surtout dans l'expression faciale, et une autre classe avec deux vues prises dans des conditions d'illumination drastiquement différentes. Néanmoins, ces mesures ne suffisent pas à nous prémunir du fait que les matrices de variance soient significativement différentes. Un certain nombre de variantes hétéroscédastiques de l'ADL ont donc été introduites ; les plus connues sont données en section 3.4.2.

Le classifieur ADL nous fournit le sous-espace linéaire des données permettant la meilleure séparation des classes. Mais, dans le contexte de la reconnaissance de visages, on peut considérer qu'une partie des variations possibles, notamment les variations dans les conditions d'illumination, sont non linéaires. Dans ce cas, l'ADL peut ne pas suffire à séparer correctement les données. L'utilisation d'une fonction de noyau peut alors permettre de définir des variantes non linéaires de l'ADL. Ces variantes sont détaillées en section 3.5.

3.3 Sous-représentation des données de visages et solutions possibles

3.3.1 Introduction

Dans le cadre de la reconnaissance de formes en général et de la reconnaissance de visages en particulier, le problème de la *sous-représentation des données* est récurrent [RJ91]. On considérera qu'il y a sous-représentation des données si le nombre d'exemples disponibles dans la base d'apprentissage est inférieur à leur dimensionnalité, ou qu'il en est trop proche. Cette notion,

ainsi que les solutions qui peuvent y être apportées, seront détaillées dans la suite de cette section.

3.3.2 Position du problème

Les performances de l'ADL standard peuvent être fortement dégradées si le nombre d'observations N est limité, comparé au nombre n de variables [Fuk90, KJMT95].

En premier lieu, la construction d'un classifieur par le biais de l'ADL standard (voir section 3.2.3) nécessite que la matrice S_w soit régulière. Puisque la matrice S_w est constituée de la somme de k matrices de variance V_j (*cf.* équation 3.2), chacune de ces matrices étant construite à partir de N_j observations de taille n, dont au plus $N_j - 1$ sont indépendantes (car elles sont centrées), la matrice de variance S_w est construite depuis au plus $\sum_{j=1}^{k}(N_j - 1) = N - k$ observations indépendantes. Par conséquent, son rang vérifie la relation :

$$\text{rang}(S_w) \quad \leq \quad \min(n, N - k) \tag{3.13}$$

Pour que la matrice S_w soit de rang plein, donc régulière, il faudrait que $N - k \geq n$; en d'autres termes il faudrait que le nombre N d'exemples disponibles soit tel que $N \geq n+k$. Cette condition n'est presque jamais vérifiée dans le contexte de la reconnaissance de visages par approches globales. Prenons l'exemple de la base ORL (*cf.* annexe A) : la taille des vecteurs-images est $n = 112 \times 92 = 10304$, et la base contient $k = 40$ personnes avec en tout $N = 400$ images de visages (dix images par classe). Pour que la matrice de variance intra-classe soit régulière, et donc l'ADL possible, il faudrait que l'on dispose non pas de 400 images de visages, mais de plus de 10344 vues. Dans la pratique, cet objectif n'est jamais atteint, et la matrice S_w est non inversible. On parlera de *problème de la singularité*. Si $N < n + k$, on qualifiera la taille de la base de *petite*, voire de *presque vide* si $N \ll n + k$.

En second lieu, il existe également un certain nombre de problèmes dans le cas où la taille N de la base approche le seuil des $n + k$. On parlera dans ce cas de bases *de taille critique*. Ces difficultés sont inhérentes à l'instabilité de l'estimation des paramètres et aux dangers de surapprentissage. Jain et Chandrasekaran [JC82] considèrent que le calcul de S_w est instable si N n'est pas au moins égal à cinq à dix fois $(n + k)$. De plus, comme nous le montrerons dans la suite de cette section, l'ADL peut souffrir d'une mauvaise capacité de généralisation, lorsqu'elle est construite depuis une base d'apprentissage de taille critique.

Les performances d'un classifieur sont généralement évaluées par une mesure de leur *capacité de généralisation* estimée par le biais du taux de reconnaissance calculé sur une base de test disjointe de la base d'apprentissage. Étudions l'impact du rapport entre la taille de la base d'apprentissage et le nombre de caractéristiques (dimension des observations) sur les performances de l'ADL.

La figure 3.3 montre les *courbes d'apprentissage* de l'ADL standard, pour différentes tailles n de données. Ces graphes donnent les taux de généralisation en fonction de la taille de la base d'apprentissage. Les données utilisées pour construire ce graphe sont n valeurs de pixels extraites de chacune des N images de $k = 2$ chiffres (classes) issues de l'une des bases du NIST.

Étant donné qu'un classifieur ADL standard ne peut être construit que si le nombre d'exemples N est supérieur ou égal à $(n + k)$, on peut observer que les courbes partent de $N = n + 2$. Un pic dans l'erreur de généralisation est observé aux alentours de $N = n + 2$. Dans le cas où l'on retient exactement $g = k - 1$ séparateurs linéaires, ceci peut être imputé à un phénomène de surapprentissage. Pour illustrer cette idée, prenons le cas extrême d'un classifieur linéaire construit à partir d'une base d'apprentissage telle que $N = n$. Dans ce cas, tous les exemples

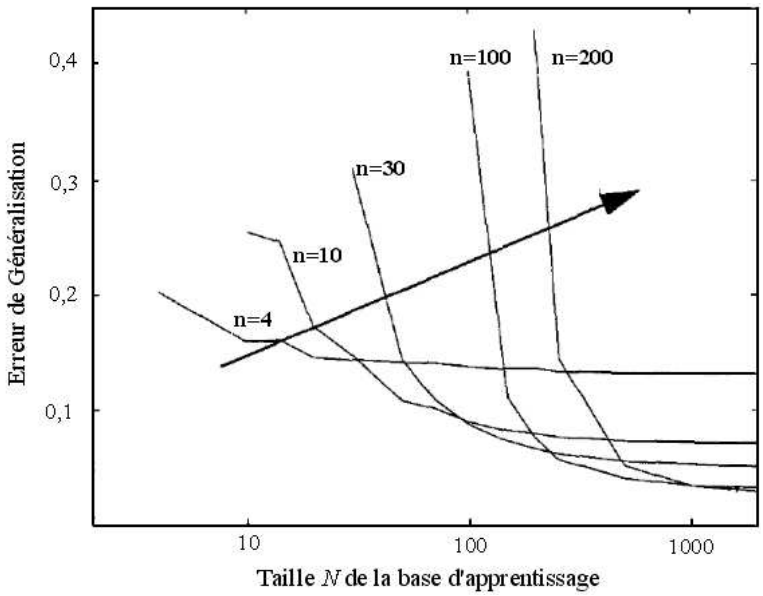


FIG. 3.3 – *Adapté de [Dui00]. Courbes d'apprentissage de l'ADL standard, pour différentes tailles n de vecteurs de caractéristiques ($n \in \{4,10,30,100,200\}$).*

d'apprentissage appartiennent à un hyperplan linéaire de dimension N. Il existe alors forcément un classifieur linéaire qui sépare parfaitement les exemples, même si ceux-ci sont aléatoirement assignés à une classe [Cov65] (en deux dimensions par exemple, cela revient à séparer deux points par une droite). Ainsi, au voisinage de la frontière $N = n + 2$, le classifieur absorbe une trop grande part du bruit de la base d'apprentissage, ce qui détériore sa capacité de généralisation.

Puis, au fur et à mesure que N augmente, l'erreur de généralisation diminue. Durant cette phase, le facteur discriminant est estimé de plus en plus précisément grâce à un nombre croissant de données. Puis, pour une taille n fixée, l'erreur de généralisation atteint son minimum en $N = \alpha(n + k)$ [JC82]. En général, la valeur de α varie entre 5 et 10. On peut déduire de cette figure que, si le nombre N d'exemples est limité, la taille de ces exemples doit être petite ; par contre, plus la taille de la base d'apprentissage augmente, plus le nombre de caractéristiques que l'on peut se permettre de conserver est important. Asymptotiquement, un nombre important d'observations garantit de meilleurs résultats.

Les performances de l'ADL sont donc très dépendantes du rapport entre le nombre N d'exemples et la taille n de ceux-ci.

De nombreuses méthodes, basées sur l'ADL et conçues pour surmonter le problème de la singularité, ont été introduites. La figure 3.4 donne un aperçu de ces techniques. Les deux premières solutions visent à augmenter artificiellement le ratio entre le nombre N d'exemples disponibles et leur taille n. Une première possibilité est d'augmenter N, pour n étant fixé. Pour cela, on peut injecter de nouveaux exemples, issus de modifications digitales des images existantes, dans la base d'apprentissage (*cf.* section 3.3.3.1). Une deuxième solution est d'appliquer une phase préalable de réduction de dimension, avant de mettre en œuvre une ADL standard sur les caractéristiques ainsi extraites. Cette technique est détaillée en section 3.3.3.2. D'autres méthodes, basées pour la plupart sur la maximisation du critère de Fisher sous des hypothèses restrictives, sont passées en revue en section 3.3.3.3. Ces techniques forment une famille désignée par le nom d'*ADL sous-optimale*. Une dernière option est de modifier légèrement le critère de Fisher, de ma-

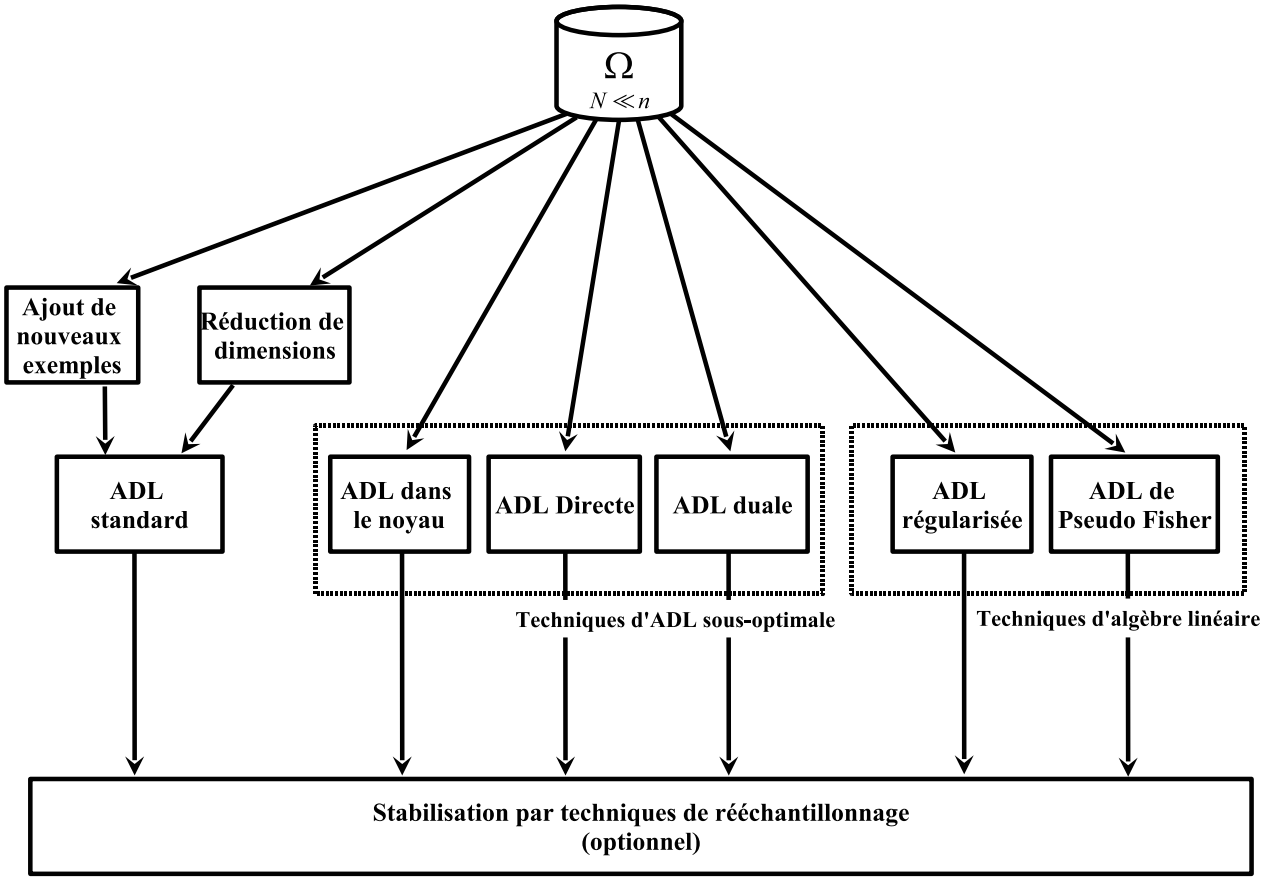


FIG. 3.4 – *Les principales techniques, issues de l'ADL, et conçues pour surmonter les difficultés inhérentes à la sous-représentation des données.*

nière à contourner le problème de l'inversion numérique de la matrice S_w, tout en garantissant une bonne séparabilité des données. Ces méthodes seront détaillées en section 3.3.3.4.

Les classifieurs ainsi construits peuvent néanmoins souffrir d'instabilité, comme nous le montrerons en section 3.3.4. Si c'est le cas, des méthodes de rééchantillonnage peuvent être mises en œuvre (voir figure 3.4).

3.3.3 Solutions possibles au problème de la singularité.

3.3.3.1 Ajout de nouveaux exemples

Li *et al.* [LL99] ont proposé un modèle linéaire, baptisé *Feature Line*, permettant de créer virtuellement, à partir de deux vues de la même personne sous des conditions différentes (illumination, expression faciale), un nombre infini de variantes de ces vues. Néanmoins, cette technique est susceptible d'introduire du bruit dans la base d'apprentissage.

Huang *et al.* [HYCL03] ont plus récemment proposé d'appliquer sur les vues disponibles des rotations et des translations. Cette technique ajoute inévitablement du bruit à la base d'apprentissage lorsque celle-ci est composée de visages normalisés (voir section 1.6), ce qui est généralement le cas.

Une approche alternative, permettant de doubler la taille de la base d'apprentissage, est

d'ajouter des versions « miroir » de chaque vue dans la base d'apprentissage [LPV05]. Une vue « miroir » est obtenue en appliquant une symétrie axiale selon l'axe vertical et central du visage. Si l'on considère que tous les visages sont symétriques suivant cet axe, les vues « miroir » ne devraient introduire que peu de bruit dans la base d'apprentissage. Cependant, les vues ainsi ajoutées n'apportent que peu de variations à l'intérieur d'une même classe, ce qui peut engendrer une capacité de généralisation moindre que si le modèle avait été construit depuis un ensemble de vues réelles en même nombre.

Dans la pratique, il n'est en général pas possible d'ajouter suffisamment d'exemples pour se départir du problème de la singularité. Considérons à nouveau l'exemple de la base ORL (*cf.* annexe A). La base d'apprentissage contient au plus $N = 400$ vecteurs-images de longueur $n = 10344$. Il nous faudrait, pour surmonter le problème de la singularité, générer plus de $10304 + 40 = 10344$ images à partir d'au plus 400 vues. Cela est inenvisageable. De telles techniques peuvent par contre être utilisées pour augmenter artificiellement le rang de la matrice S_w (ce qui peut améliorer les performances de certains classifieurs). Cependant, elles sont pour la plupart susceptibles d'introduire du bruit dans la base d'apprentissage.

3.3.3.2 Utilisation d'une phase préliminaire de réduction de dimension

Les techniques mettant en œuvre une réduction de dimensions en amont de l'ADL visent à réduire n, tout en laissant N inchangé. L'Analyse en Composantes Principales est une technique souvent utilisée dans ce but, puisqu'elle permet de diminuer les dimensions du problème de n à $M \ll n$, tout en retenant les M variables (combinaisons linéaires des variables en entrée) non corrélées et expliquant la majeure partie de la dispersion des données.

La technique des *fisherfaces* La méthode des *fisherfaces* [BHK97], brièvement introduite en section 2.2.2.4, est certainement la plus connue de ces approches. Détaillons le processus. Si l'on se base sur l'algorithme de résolution de l'ADL de Fukunaga [Fuk90], détaillé en section E.3.2 (p. 194) de l'annexe E, la méthode des *fisherfaces* compte trois étapes : la première consiste à effectuer une ACP dans l'espace original des visages (en retenant M axes principaux), la seconde à blanchir la matrice de covariance intra-classe S'_w des observations projetées dans le sous-espace principal ainsi défini, et enfin la troisième étape est de construire une ACP sur les centres des classes, dans l'espace propre blanchi de S'_w. On en déduit la matrice de projection W de l'ADL globale, contenant les g facteurs discriminants.

La technique des *fisherfaces* ainsi obtenue est dépendante de deux paramètres M et g que Belhumeur *et al.* [BHK97] choisissent de fixer à $M = N - k$ (valeur maximale de M telle que S_w soit de rang plein), et $g = k - 1$. Cela revient à appliquer une ADL standard sur N exemples de taille $M = N - k$, c'est-à-dire sur une base de *taille critique*. Dans ce contexte, il y a de fortes chances que le classifieur ADL souffre de surapprentissage (*cf.* section 3.3.2), et que ses performances en soient affectées. La capacité de généralisation des *fisherfaces* pourrait donc être améliorée, par exemple en réduisant la taille M des vecteurs en entrée.

Swets et Weng ont présenté une technique similaire dans [SW96], mais basée sur l'utilisation de $M < N - k$ axes principaux. Ils choisissent la valeur du paramètre M à l'aide d'une mesure de l'*énergie de dimension* (voir p. 30 de la section 2.2.2.1), à un seuil de 5%. Le paramètre M est donc choisi de manière à ce que le sous-espace principal explique 95% de la variance des vecteurs-images, mais sans tenir compte des besoins de l'ADL en aval. Une technique équivalente est mise en œuvre pour choisir g. Cette technique est souvent, par abus de langage, désignée par le terme de *fisherfaces*.

Modèles de Fisher Améliorés (MFA) Liu et Wechsler [LW98, LW00] proposent deux algorithmes dits de Fisher Améliorés, notés par la suite MFA-1 et MFA-2.

L'apport de l'algorithme MFA-1 [LW98] sur la technique de Swets et Weng est qu'il prend également en compte les besoins de l'ADL en aval dans le choix du paramètre M. Leur technique, graphique, est basée sur l'étude conjointe de l'éboulis des valeurs propres de la matrice S_T, et de l'éboulis des valeurs propres de la matrice S'_w (un exemple d'éboulis des valeurs propres est donné en figure 2.3 du chapitre précédent). On choisit un certain nombre de valeurs-candidates M_i pour le paramètre M, dans la région d'inflexion de l'éboulis de S_T. Pour chaque candidat M_i, on trace l'éboulis des valeurs propres de la matrice S'_w correspondante (c'est-à-dire construite dans l'espace de l'ACP à M_i axes principaux), et on retient le M_i permettant d'obtenir des valeurs propres de S'_w qui ne soient pas trop faibles. Cette technique de sélection du M optimal est coûteuse et très empirique, donc difficilement utilisable. De plus, elle est basée sur l'hypothèse que les vecteurs propres associés aux plus petites valeurs propres de S'_w encoderaient du bruit, ce qui n'a pas été démontré.

Alternativement, Liu et Wechsler [LW98] (algorithme MFA-2) proposent d'effectuer l'ACP préalable avec $M = N - k$, mais par contre de ne retenir que les $M' < N - k$ vecteurs propres de S'_w associés aux plus grandes valeurs propres dans la deuxième étape de l'algorithme. La suite du processus reste la même que pour les *fisherfaces*. La stratégie de choix du paramètre M', basée sur l'étude de l'éboulis des valeurs propres de S'_w, reste très empirique.

Avec un nombre g suffisant de vecteurs de projection, les deux algorithmes (MFA-1 et MFA-2) donnent des taux de reconnaissance comparables sur une sous-base de la base FERET (voir annexe A). Les résultats obtenus sont meilleurs que ceux de la technique des *fisherfaces* [LW98]. Liu et Wechsler [LW98] mettent également en évidence l'existence d'une valeur optimale $g^* \leq k - 1$ du paramètre g, qui permet d'obtenir le meilleur taux de reconnaissance sur une base de test indépendante de la base d'apprentissage : pour $g < g^*$ les taux de reconnaissance augmentent lorsque g augmente, puis passé g^* les taux de reconnaissance stagnent, voire diminuent. Cependant, ils n'ont pas proposé dans leurs travaux d'heuristique pour déterminer cette valeur optimale.

Discussion Les résultats expérimentaux obtenus par différents auteurs [SW96, BHK97, Zha99, PMRR00, GSC01, Yan02] ont montré qu'en général, la technique basée sur une ACP, suivie d'une ADL, (notée ACP+ADL) est plus performante que la méthode des *eigenfaces*, basée sur une simple ACP (*cf.* section 2.2.2.1). Ce résultat semble logique dans la mesure où, pour les *eigenfaces*, le critère à minimiser est l'erreur Euclidienne de reconstruction, moins adapté à la classification des données que le critère de Fisher, lié à la séparabilité des données. Mais la phase préliminaire d'ACP peut conduire au rejet d'information potentiellement discriminante [CLK+00, YY01, BLP02, DY03, HLLM02, YZY03, WT04a] et conserver en revanche de l'information non discriminante, c'est-à-dire du bruit. Les composantes sélectionnées par l'ADL étant des combinaisons linéaires (en nombre inférieur) de ces variables bruitées, la phase préliminaire d'ACP peut nuire à la capacité de généralisation de l'ADL. En particulier, une partie du noyau de la matrice de variance intra-classe S_w des données originales peut être rejetée par la phase préliminaire d'ACP. Or, comme nous le verrons ci-après, l'information provenant du noyau de S_w peut avoir un fort pouvoir discriminant. Dans le but de conserver tout ou partie de cette information, des techniques dites *d'ADL sous-optimale* ont été introduites.

3.3.3.3 L'ADL sous-optimale

Chen *et al.* [CLK+00] ont montré que le noyau de S_w peut contenir de l'information discriminante, si la projection de S_b est non nulle dans les directions ainsi définies. Notons W une

telle direction. Par définition, on a :

$$W^T S_w W = 0 \quad \text{et} \quad W^T S_b W \geq 0 \tag{3.14}$$

ce qui implique que le critère de Fisher donné en équation (3.11) atteint un maximum global dans cette direction :

$$J(W) = \frac{|W^T S_b W|}{|W^T S_w W|} = +\infty$$

Un certain nombre de techniques permettant de conserver de l'information issue du noyau de S_w ont été introduites ; certaines ne conservent que l'information provenant du noyau (techniques d'ADL dans le noyau), tandis que d'autres (ADL Directe et ADL Duale) conservent une partie de l'information provenant de l'extérieur du noyau de S_w. Dans cette partie, nous ne définirons que superficiellement les principales techniques d'ADL sous-optimale. Plus de détails concernant ces techniques sont données en annexe F (p. 197).

L'ADL dans le noyau La première technique d'ADL dans le noyau a été présentée en 2000 par Chen *et al.* dans [CLK+00] et est notée ADL_0. Elle consiste à extraire une base orthonormée du noyau de S_w, puis à projeter les centres des classes dans ce noyau avant de mettre en œuvre une ACP sur ces centres projetés. Cela revient à remplacer l'étape 1. de blanchiment de l'algorithme de Fukunaga (p. 194 de l'annexe E) par une étape d'extraction du noyau (dont une base est constituée des vecteurs propres associés aux valeurs propres nulles), et de projection des centres sur ce noyau. Évidemment, si la matrice S_w est de rang plein, alors cette technique ne peut être appliquée et c'est une ADL standard qui est utilisée.

Cette technique nécessite le calcul du rang de la matrice S_w, qui est une opération mal posée. De plus, étant données les très grandes dimensions des données, la diagonalisation de la matrice S_w est très coûteuse et instable, et le calcul des derniers vecteurs propres est généralement imprécis. C'est pourquoi Chen *et al.* mettent en œuvre, en amont de l'ADL_0, une technique d'agglomération de pixels leur permettant de réduire la dimensionnalité du problème, mais pouvant conduire à une perte d'information discriminante. Des expériences menées par Cevikalp *et al.* [CNWB05] mettent en évidence le fait que, plus la taille du noyau de S_w est importante, plus l'ADL_0 est performante. Par conséquent, toute phase préalable de réduction de dimension (engendrant une réduction de la taille du noyau) devrait être évitée.

Afin de réduire la complexité de la première étape d'extraction du noyau de l'ADL_0, Huang *et al.* [HLLM02] proposent de rejeter en amont l'information provenant à la fois du noyau de S_w et du noyau de S_b, car cela n'a aucun effet sur le critère de Fisher (3.11) (*cf.* section F.1.2 de l'annexe E). Huang *et al.* montrent que $\text{Ker}(S_T) = \text{Ker}(S_w) \cap \text{Ker}(S_b)$ et, partant de ce constat, ils proposent de mettre en œuvre une phase d'ACP préalablement à l'ADL_0, ne retenant dans le sous-espace principal que les vecteurs propres associés à des valeurs propres non nulles (neutralisation du noyau de S_T). Cette technique est désignée par le sigle ACP+ADL_0. Elle présente le désavantage que la phase préliminaire d'ACP engendre un coût calculatoire supplémentaire.

La technique des Vecteurs Discriminants Communs (VDC), introduite dans le contexte de la reconnaissance de visages par Cevikalp *et al.* [CNWB05], vise à extraire un *vecteur discriminant commun* à tous les membres d'une même classe, en travaillant dans l'espace nul de la variance intra-classe. Il en résulte deux algorithmes très efficaces en temps de calcul. Pour plus de détails, se référer à la section F.1.3 de l'annexe F.

Toutes ces techniques utilisent pleinement l'information provenant du noyau de S_w, maximisant ainsi définitivement le critère de Fisher (3.11). Par contre, elles ne prennent en compte aucune information hors du noyau de S_w, dont l'ajout n'a aucun effet sur ce critère. Néanmoins,

pour évaluer un classifieur ADL, on préférera souvent une estimation de la capacité de généralisation sur des bases de test indépendantes à la valeur du critère de Fisher, car ce dernier est moins lié aux résultats de classification. Or, des résultats expérimentaux, montrent qu'il pourrait exister de l'information discriminante (au sens des taux de classification) en dehors du noyau de S_w [YY01]. De plus, les performances de la plupart de ces techniques diminuent lorsque la taille du noyau diminue, c.-à-d. lorsque le nombre d'exemples augmente ou que la dimension des observations diminue [CNWB05].

L'ADL Directe Les techniques d'ADL dans le noyau rejettent systématiquement toute information hors du noyau de S_w. De plus, ces méthodes nécessitent généralement de déterminer et/ou de manipuler le noyau de la matrice S_w, difficilement évaluable. Pour pallier ces inconvénients, Yu et Yang proposent dans [YY01] la technique d'ADL Directe (ADLD), qui consiste à inverser l'ordre des diagonalisations : on commence par blanchir S_b, avant de minimiser la variance intra-classe des données blanchies. Lors de cette dernière phase de minimisation, on peut facultativement ne garder que le noyau de la matrice de variance intra-classe projetée. Pour plus de détails, se référer à la section F.2 de l'annexe F. Les résultats expérimentaux sont significativement meilleurs sans la phase optionnelle de rejet de l'information hors du noyau qu'avec. Cela montre qu'il peut exister de l'information discriminante hors du noyau de S_w.

L'avantage principal de la technique d'ADL Directe sur celles d'ADL dans le noyau est qu'elle peut tirer parti de l'information hors du noyau de S_w. De plus, l'ADL Directe, comme son nom l'indique, peut être appliquée directement sur les vecteurs-images de visages, sans nécessiter d'étape préliminaire de réduction de dimension (agglomération de pixels ou ACP). Par contre, Yu et Yang font l'hypothèse que le noyau de S_b ne contient aucune information discriminante, ce qui est faux en général et nuit aux performances de leur technique (en comparaison avec la méthode d'ADL_0), comme le montrent les résultats d'évaluation détaillés dans le tableau 3.3 et analysés en section 3.3.3.5. Ceux-ci mettent en évidence le fait que l'ADL_0 est généralement plus performante que l'ADL Directe, qui elle-même peut ou non être meilleure que la technique des *fisherfaces*, suivant les bases considérées.

L'ADL Duale Nous avons vu qu'il existe de l'information discriminante à la fois dans le noyau de S_w et hors de celui-ci. L'ADL standard est effectuée dans l'espace image de S_w, tandis que l'ADL_0 est construite depuis le noyau de S_w. Récemment, des techniques permettant de prendre en compte conjointement de l'information provenant de ces deux sous-espaces ont vu le jour.

Wang et Tang ont introduit dans [WT04a] une méthode dite d'ADL Duale. Cette technique consiste à extraire le sous-espace principal $\mathcal{W}$ de S_w et à effectuer une ADL standard dans cet espace. Dans un même temps, on procède à une ADL_0 sur le complémentaire $\mathcal{W}^\perp$ du sous-espace principal. Lorsqu'un nouvel individu se présente, on calcule sa distance à chacune des classes en tenant compte à la fois de la distance dans $\mathcal{W}$ et son complémentaire, par le biais d'une mesure de dissimilarité inspirée de celle introduite par Moghaddam et Pentland en 1997 [MP97] dans le cadre des sous-espaces probabilistes (*cf.* section 2.2.2.2, p. 30). L'algorithme, détaillé en section F.3 de l'annexe F, est très coûteux. De plus, il repose sur l'estimation de la valeur propre moyenne sur $\mathcal{W}^\perp$ par l'ajustement d'une fonction non linéaire sur le spectre des valeurs propres disponibles. Or, cette estimation potentiellement imprécise a une grande influence sur la mesure de distance retenue pour la classification.

Yang *et al.* proposent dans [YZY03] une autre technique d'ADL Duale. On commence par neutraliser le noyau de S_T, à la manière de Huang *et al.* [HLLM02]. Puis, on diagonalise la matrice de variance projetée S'_w, de manière à déterminer le noyau $\mathcal{W}'^\perp$ de S'_w, et on applique

une ADL_0 dans $\mathcal{W}'^{\perp}$. Si plus de vecteurs de projection sont nécessaires à une bonne classification, on pourra ajouter des vecteurs propres issus d'une ADL standard, menée en parallèle sur l'espace-image $\mathcal{W}'$ de la matrice S'_w. Plus de détails concernant cette technique sont donnés en section F.3 de l'annexe F. Cette méthode est séduisante, en ce sens qu'elle permet, si nécessaire, de prendre en compte de l'information hors du noyau de S_w. Cependant, elle est très coûteuse en termes de temps de calcul (tant pour la construction du modèle que pour la phase de classification) et la règle permettant de décider d'ajouter plus de vecteurs propres n'est pas claire.

3.3.3.4 Les techniques d'ADL modifiées

Ces méthodes visent, par une légère modification du critère de Fisher et l'utilisation de techniques d'algèbre linéaire, à contourner le problème de la singularité.

L'ADL régularisée Les techniques dites d'*ADL Régularisées* ajoutent pour la plupart une matrice diagonale de perturbation à la matrice de covariance intra-classe, de manière à la rendre inversible. Considérons la régularisation suivante [Fri89] : $S_w + \sigma I_n$ où n est la taille des observations, I_n est la matrice identité de taille $n \times n$, et σ l'élément perturbateur. Il est facile de vérifier que $S_w + \sigma I_n$ est définie positive, et donc régulière. Une valeur de σ trop grande conduit à une perte d'information discriminante, tandis que si l'on choisit un σ trop petit, l'inversion est instable [SD96]. La valeur optimale de σ est très dépendante de la base choisie, et est communément déterminée en utilisant une stratégie de validation croisée [KJMT95], ce qui est très coûteux. Plus récemment, de nouvelles techniques de régularisation ont été introduites [ZH01, DY03, LPV05] ; toutes demandent un ajustement coûteux de leurs paramètres.

L'ADL de Pseudo-Fisher (ADL^+) Une autre voie, sans ajout de paramètres par rapport à une ADL standard, est de remplacer l'inverse S_w^{-1} de la matrice de dispersion S_w par sa pseudo-inverse S_w^+ dans l'algorithme de résolution standard [TBGL86]. Cette technique, que nous désignerons par l'acronyme ADL^+, est souvent appelée *Analyse Discriminante Linéaire de Pseudo Fisher* [Dui95, Fuk90, SD96, RD98, SD99, Dui00]. La matrice W de projection de l'ADL est constituée des vecteurs propres de $S_w^+ S_b$ associés aux plus grandes valeurs propres.

Définition 3.1 *Notons r le rang de S_w. Si $r < n$, le calcul de S_w^+ passe par la technique de Décomposition en Valeurs Singulières (DVS) de la matrice S_w, définie comme suit : $S_w = U\Sigma V^T$, où U et V sont deux matrices orthonormées de taille $n \times n$, et $\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r, 0, \ldots, 0)$ est la matrice diagonale contenant les valeurs singulières. Si l'on note $\Sigma^+ = \text{diag}(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_r}, 0, \ldots, 0)$, alors la pseudo-inverse S_w^+ (aussi appelée pseudo-inverse de Moore-Penrose) de S_w s'écrit :*

$$S_w^+ = V\Sigma^+ U^T \tag{3.15}$$

Propriété 3.1 *Si $r = n$, la matrice S_w est inversible et $S_w^+ = S_w^{-1}$. Sinon, si $r < n$, on peut montrer que S_w^+ est la meilleure approximation de S_w^{-1} au sens des moindres carrés :*

$$S_w^+ = \underset{X \in \mathbb{R}^{n \times n}}{\text{Argmin}} \|S_w X - I_n\|_2 \tag{3.16}$$

où I_n est la matrice identité de taille $n \times n$.

La propriété 3.1 implique que, si les données ne souffrent pas du problème de la singularité, mettre en œuvre une ADL^+ revient à appliquer une ADL standard. Sinon, si la matrice S_w est singulière, l'ADL^+ revient à définir les r premiers facteurs discriminants par une ADL dans l'espace image

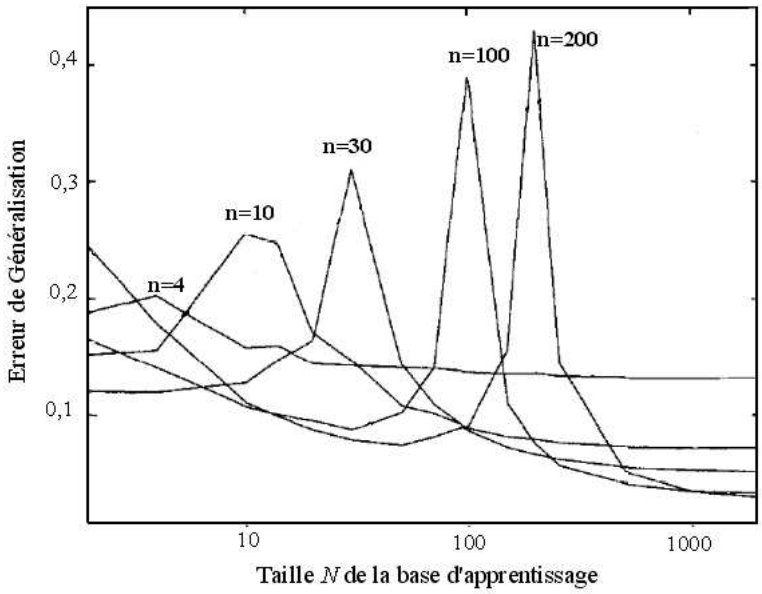


FIG. 3.5 – *Adapté de [Dui00]. Courbes d'apprentissage de l'ADL$^+$, pour différentes tailles n de vecteurs de caractéristiques ($n \in \{4,10,30,100,200\}$).*

de S_w, puis à compléter l'espace discriminant par une ACP sur les vecteurs orthogonaux aux r premiers axes discriminants. Les travaux de Gower [Gow66] ont montré que, dans les conditions de singularité de la matrice de dispersion intra-classe S_w, utiliser la pseudo-inverse revient à mettre en œuvre une distance de Mahalanobis généralisée.

Étudions l'impact sur les performances de l'ADL$^+$ du rapport entre le nombre d'exemples et leur taille. La figure 3.5, extraite de [Dui00], donne les courbes d'apprentissage de l'ADL$^+$, pour différentes tailles des vecteurs de caractéristiques. Ces courbes sont tracées selon la même procédure que pour l'ADL standard (voir section 3.3.2). On peut vérifier que, pour $N > n + 2$, le comportement de l'ADL$^+$ est le même que celui de l'ADL standard (voir figure 3.3), puisque dans ce cas les deux techniques sont strictement équivalentes. Pour plus de détails concernant cette portion de courbe, on peut donc se référer à la section 3.3.2. Pour N allant de 4 à $n + 2$, l'erreur de généralisation commence généralement par baisser, pour atteindre un minimum local, avant d'augmenter et d'atteindre son maximum en $N = n + 2$. Le minimum global n'est atteint que pour $N \gg n$ (partie droite de la courbe). La figure 3.5 montre que, dans le cas d'espaces presque vides (partie à l'extrême gauche de la courbe), plus le nombre de caractéristiques augmente, plus l'erreur de généralisation augmente. Raudys et Duin [RD98] donnent, sous certaines hypothèses, une expression asymptotique de l'erreur de généralisation de l'ADL$^+$, expliquant théoriquement ce comportement. On peut considérer que, dans le cas d'espaces presque vides, le nombre d'exemples disponibles ne permet pas d'estimer précisément les facteurs principaux. Aussi l'ADL$^+$ n'est-elle pas adaptée à des espaces presque vides [RD98]. Afin de pallier ce problème, on peut néanmoins augmenter artificiellement le ratio $\frac{N}{n}$, soit en introduisant de nouveaux exemples, soit en utilisant une phase préliminaire de réduction de dimension telle qu'une ACP.

Dans [HJP03, HP04], Howland *et al.* introduisent une version généralisée de l'ADL, permettant de contourner le problème de la singularité, par le biais d'une Décomposition en Valeurs Singulières Généralisée. Ye *et al.* [YJPP04] ont montré que la solution obtenue par ADL Généralisée (ADLG) est la même que la solution obtenue par une ADL$^+$ usuelle. L'algorithme exact proposé étant très coûteux, ils introduisent alternativement un algorithme dit approxima-

Rang	Performance	Coût de construction	Coût de classification	Coût de stockage
1	VDC, ACP+ADL$_0$	ADLD	VDC,ACP+ADLD$_0$	VDC,ACP+ADLD$_0$
2	*fisherfaces*	VDC	*fisherfaces*, ADLD	*eigenfaces*, ADLD
3	ADLD	*eigenfaces*	*eigenfaces*	*fisherfaces*
4	*eigenfaces*	*fisherfaces*		
5		ACP+ADLD$_0$		

TAB. 3.2 – *Adapté de [CNWB05]. Classement des principales méthodes présentées dans cette section, de la plus efficace à la moins efficace, en fonction de quatre critères. Le premier mesure la performance de l'algorithme (erreur de généralisation moyenne sur une base de test indépendante de la base d'apprentissage). Les second et troisième critères correspondent respectivement au coût de la construction du modèle et de la classification, en termes de nombres d'opérations du système. Le quatrième critère mesure le coût du stockage du modèle. Les méthodes testées sont : les Vecteurs Discriminants Communs (VDC), l'algorithme d'ADL dans le noyau de Huang et al. [HLLM02] (ACP+ADL$_0$), l'ADL Directe (ADLD) de Yu et Yang [YY01], ainsi que les techniques des eigenfaces et des fisherfaces.*

tif, où les classes ne sont pas représentées par l'ensemble de leurs points, mais par un ensemble de centres de sous-groupes. Ces sous-groupes sont déterminés automatiquement par la mise en œuvre d'un *algorithme des K-moyennes* [JD88] à l'intérieur de chaque classe. Les résultats expérimentaux montrent de bons résultats ; néanmoins les performances sont très dépendantes du choix du nombre de sous-groupes par classe dans l'algorithme de K-moyennes.

3.3.3.5 Comparaison des performances des différentes techniques

Le tableau 3.2, adapté de [CNWB05], donne un classement au regard de divers critères de quelques-unes des techniques d'ADL conçues pour surmonter le problème de la singularité et présentées ci-avant. Le classement des performances est basé sur les taux de généralisation obtenus sur les bases AR et Yale A (*cf.* annexe A).

On remarque que les techniques les plus performantes en termes de taux de classification sont les techniques de l'ADL$_0$, et les VDC. Il est à noter que l'ADLD n'arrive qu'en troisième position, entre les *fisherfaces* et les *eigenfaces*.

Le tableau 3.3, regroupe les résultats d'expérimentations de plusieurs auteurs sous la forme d'un classement de la plupart des méthodes présentées dans cette section. Les méthodes testées sont, en plus des techniques évaluées dans le cadre du tableau 3.2 : les Modèles de Fisher Améliorés (MFA) de Liu *et al.* [LW01], l'ADL dans le noyau de Chen *et al.* [CLK$^+$00] (ADL$_0$), l'ADL Duale de Yang *et al.* [YZY03], les ADL Régularisées de Friedman [Fri89] (ADLRF) et de Dai et Yuen [DY03] (ADLRD), ainsi que la technique d'ADL Généralisée (ADLG) de Howland *et al.* [HJP03] (liée à la technique de la pseudo-inverse).

Source	Base	N	k	n	Conclusion
[PPP04]	Yale	164	15	non précisé	ADLG>ADL Duale >ADLRF=ADL_0 >ADLD
[DY03]	ORL	200	40	2576	ADLRD>ADLD>*fisherfaces*
[LW01]	FERET	738	369	6144	MFA-2≃MFA-1>*fisherfaces*
[DQJ04]	ORL	200	40	10304	MFA>ADLD>*fisherfaces*
[HLLM02]	ORL	80–360	40	10304	ACP+ADL_0 >ADLD>*fisherfaces*
	FERET	140–350	70	10304	ACP+ADL_0 >ADLD>*fisherfaces*
[CNWB05]	AR	350	50	66378	ACP+ADL_0 ≃VDC>*fisherfaces*≥ADLD
	Yale A	150	15	19152	ACP+ADL_0 ≃VDC>*fisherfaces*>ADLD

TAB. 3.3 – *Comparaison des performances des algorithmes proposés, basés sur les résultats d'expérimentations menés par différents auteurs sur diverses bases de visages et en fonction du nombre k de classes, du nombre N d'exemples disponibles, et de leur taille n. Le symbole '≃' signifie que les performances des deux techniques sont comparables. L'opérateur '≥' désigne les cas où la première méthode est meilleure que la seconde, mais de manière non significative ; on utilisera le symbole '>' si la différence de performance est significative.*

Il faut noter que, dans [DQJ04] et [CNWB05], la méthode des *eigenfaces* (voir section 2.2.2.1) est également testée, et est significativement moins performante que les méthodes basées sur l'ADL.

La première conclusion que l'on peut tirer du tableau 3.3 est que les performances relatives des algorithmes varient en fonction des bases utilisées, et du ratio nombre d'exemple / dimensionnalité du problème. L'exemple le plus frappant est celui de l'ADL Directe (ADLD), qui semble être peu performante sur les bases presque vides [CNWB05] (en comparaison avec les autres méthodes), tandis que ses résultats sont meilleurs sur des bases de plus grande taille [DY03]. La méthode des *fisherfaces* est généralement parmi les moins performantes mais dans certains cas elle peut être meilleure que la technique d'ADL Directe. L'amélioration de Liu *et al.* semble être efficace, puisque les Modèles de Fisher Améliorés (MFA) fonctionnent mieux que les *fisherfaces* [LW01, DQJ04], et peuvent même être plus performants que l'ADL Directe [DQJ04].

Les techniques les plus stables (dont les performances varient le moins en fonction des caractéristiques de la base utilisée) sont les algorithmes d'ADL dans le noyau (ADL_0 et ACP+ADL_0), les Vecteurs Discriminants Communs (VDC), et les techniques d'ADL régularisées (ADLRF et ADLRD). L'ADL Duale, ainsi que l'ADL Généralisée, bien que testées sur un nombre moins important de bases, semblent également très performantes.

3.3.4 Stabilisation par rééchantillonnage

Quand les données sont de très grandes dimensions et en nombre faible comparé à ces dimensions, il peut être difficile de construire un unique classifieur qui soit performant. Généralement, les classifieurs linéaires construits à partir de telles bases d'apprentissage sont *biaisés* et, étant donné que leurs facteurs discriminants sont estimés de manière imprécise, ils présentent une *variance* importante car l'estimation de leurs paramètres (coefficients) est mauvaise. De plus,

bien souvent, de tels classifieurs sont *instables* : des changements dans la base d'apprentissage peuvent causer des changements importants dans les vecteurs de projection du classifieur et donc des différences dans les performances (l'étude du tableau 3.3 illustre bien ce phénomène). Une solution pour améliorer les performances de ces classifieurs est d'en construire plusieurs au lieu d'un seul, et de les combiner pour donner naissance à un unique modèle, plus performant. C'est à cette fin qu'ont été introduites les techniques de *bagging* [Bre96], de *boosting* [FS96] et des *sous-espaces aléatoires* [Ho98]. Ces trois méthodes, initialement conçues pour stabiliser ou améliorer les performances des arbres de décision, peuvent également être utilisées avec succès pour les classifieurs ADL, et notamment l'ADL de pseudo Fisher (ADL^+) [SD02]. Ces méthodes, ainsi que leurs principales applications dans le cadre de la reconnaissance de visages, sont passées en revue en annexe G (p. 205).

L'utilisation de techniques de rééchantillonnage pour l'ADL, dans le contexte de la reconnaissance de visages, est émergente. À notre connaissance, peu de méthodes ont été proposées à ce jour ; celles-ci illustrent cependant le fait qu'il existe deux manières de mettre en œuvre le rééchantillonnage : directement sur les données en entrée, ou bien sur un ensemble de caractéristiques qui en sont extraites (par le biais d'une ACP par exemple).

Dans l'espace initial des données, les auteurs ont appliqué essentiellement des techniques proche du *bagging*, tout en cherchant à favoriser un nombre égal d'observations dans chaque classe, soit en rééchantillonnant uniformément à l'intérieur des classes (technique de Lu et Jain [LJ03]), soit en supprimant un certain nombre de classes [WT04b]. La technique des *sous-espaces aléatoires* n'a, à notre connaissance, jamais été utilisée sur les données en entrée, bien que cela soit le cas -avec succès- pour l'ACI [CLLC04].

Dans l'espace des caractéristiques issues de l'ACP, Wang et Tang appliquent avec succès la méthode des sous-espaces aléatoires, couplée à une ADL_0.

Les possibilités de couplage entre variantes de l'ADL et techniques de rééchantillonnage sont très nombreuses. Encore faut-il choisir la technique de rééchantillonnage la plus adaptée au classifieur, et concevoir un schéma de mise en œuvre qui soit le moins coûteux possible.

3.3.5 Conclusion

Dans cette section, nous avons mis en évidence le fait qu'il est impossible d'utiliser directement l'ADL sur les visages modélisés de manière 1D (les images sont représentées par le biais de vecteurs), à cause du problème de la singularité. De nombreux auteurs se sont penchés sur ce problème, et ont conçu des techniques visant à le surmonter. Parmi ces techniques, on compte l'ajout de nouveaux exemples, l'utilisation d'une phase préliminaire de réduction de dimension, l'ADL sous-optimale et les techniques basées sur la modification du critère de Fisher. Nous avons vu que, selon les caractéristiques des bases de visages utilisées pour l'évaluation, les performances relatives de ces techniques varient. Afin de stabiliser le processus, nous avons présenté les diverses techniques de rééchantillonnage, qui sont à ce jour émergentes dans le contexte de la reconnaissance de visages.

3.4 Variantes de l'ADL dans le cas où ses hypothèses sont non vérifiées

3.4.1 Introduction

Les hypothèses de multinormalité et d'homoscédasticité, qui seules peuvent garantir que le classifieur ADL soit optimal au sens de la règle de Bayes, sont difficilement vérifiables dans

le contexte de la reconnaissance de visages, à cause des très grandes dimensions des données. L'ADL étant relativement robuste à des écarts à la condition de multinormalité (voir section 3.2), la plupart des techniques supposent cette hypothèse vérifiée. Par contre, il existe de nombreuses variantes de l'ADL adaptées au cas où les données ne vérifient pas l'hypothèse d'homoscédasticité (on parle alors d'hétéroscédasticité). Ces variantes de l'ADL seront passées en revue en section 3.4.2. Nous avons également évoqué en section 3.2.4 le fait que l'ADL est sensible à la présence d'observations aberrantes ; les méthodes présentées en section 3.4.3 sont conçues pour être moins influencées par de telles observations.

3.4.2 L'ADL hétéroscédastique

Nous avons évoqué en section 3.2.3 l'Analyse Discriminante Quadratique [McL04] (pour plus de détails se référer à la section E.2.2 (p. 193) de l'annexe E), et nous avons conclu en section 3.2.4 que cette approche n'est pas adaptée aux cas où l'on ne dispose que de peu d'exemples par classe, car elle nécessite une estimation précise de chacune des matrices de variance des classes. De nombreuses autres variantes hétéroscédastiques de l'Analyse Discriminante ont été introduites. Les plus anciennes et plus connues de ces méthodes sont détaillées dans [DK82, Fuk90, McL04]. La plupart du temps, ces algorithmes sont définis pour deux classes et l'extension à plus de deux classes n'est pas directe. De plus, certaines de ces approches nécessitent une procédure d'optimisation itérative.

Parmi les méthodes non paramétriques, nous pouvons citer les techniques de Buturovic [But94] et de Liu *et al.* [LSG04]. Tout comme l'ADL, le but de ces techniques est de construire un sous-espace conservant le plus possible la séparation initiale des classes. La séparation est évaluée par une procédure au plus proche voisin, ce qui rend la construction du modèle coûteuse.

Une approche hétéroscédastique très coûteuse, basée sur la minimisation de l'erreur Bayésienne par l'utilisation d'un algorithme de recuit simulé [BJS86] et d'une intégration exacte dans le sous-espace, est introduite dans [RW99].

Une méthode plus directe, présentée dans [DM77], est basée sur la divergence de Kullback. Néanmoins elle nécessite une procédure d'optimisation itérative qui est beaucoup plus complexe que le critère de Fisher. Alternativement, Kumar et Andreou [KA98] ont introduit une technique hétéroscédastique basée sur l'utilisation du maximum de vraisemblance. Dans la technique de Hastie et Tibshirani [HT96], un modèle de mélange de gaussiennes de distributions inconnues et estimées par un critère du maximum de vraisemblance selon l'algorithme Expectation Maximization (EM) [Moo96]. La sous-représentation des données freine l'utilisation de telles techniques dans le cadre de la reconnaissance de visages.

Récemment, Loog et Duin ont proposé une autre extension hétéroscédastique de l'ADL [LD04], basée sur une redéfinition de la matrice de variance inter-classe, permettant de prendre en compte lors de son calcul des directions qui, sous l'hypothèse d'homoscédasticité, auraient été rejetées.

La plupart de ces techniques nécessitent l'estimation des matrices de variance de chacune des classes et sont très affectées par la sous-représentativité de la base d'apprentissage. De plus, tout comme l'ADL usuelle, ces techniques sont sensibles à la présence de données aberrantes.

3.4.3 L'ADL robuste

Les observations aberrantes constituent une source de contamination, déformant l'information obtenue à partir des données brutes. Afin de surmonter ce problème, deux approches sont possibles. La première consiste à exclure les données de la base d'apprentissage en amont de la construction du modèle. Il faut pour cela les détecter, c'est-à-dire connaître les éléments qui

les caractérisent, ce qui n'est pas toujours évident lorsque les dimensions du problème sont très grandes. D'autres approches consistent à les inclure dans un modèle qui soit moins sensible à leur présence [BL94]. Il s'agit de *procédures d'accommodation*, au sens où l'on essaie de construire un modèle le plus précis possible, et ce malgré la présence supposée de données aberrantes.

Intéressons-nous tout d'abord à la définition précise de ce qu'est une observation aberrante. Pour une description détaillée se référer à [Pla05]. Barnett et Lewis, en 1994, caractérisent ainsi une observation (ou un ensemble d'observations) qui semble être inconsistante avec le reste des données. Dans notre contexte, le qualificatif « aberrante » peut donc s'appliquer aussi bien à un visage qu'à une classe (qui constitue un ensemble de visages).

Les techniques permettant de rendre le classifieur ADL moins sensible aux données aberrantes isolées sont détaillées en section 3.4.3.1, tandis que celles visant à réduire l'influence des classes aberrantes sont présentées en section 3.4.3.2.

3.4.3.1 L'ADL résistante aux observations aberrantes

Les premières variantes robustes de l'ADL reposent sur l'utilisation d'estimateurs robustes pour le calcul de la moyenne et/ou de la variance. Les plus anciennes de ces techniques [AL77, RBRH78, Cam82, BT87] sont basées sur l'utilisation d'estimateurs insuffisamment précis, du fait de leur manque d'*équivariance* [11], ou d'un *point d'effondrement* [12] relativement bas. Des travaux plus récents, basés sur l'utilisation d'estimateurs ayant un point d'effondrement plus haut, sont prometteurs [CR92, HM97, HF00, CD01, HVD00]. Néanmoins, ces méthodes peuvent être trop peu performantes si les données sont sous-représentées, ce qui est généralement le cas pour la reconnaissance de visages (voir section 3.3).

Une autre technique [Pir01] repose sur l'utilisation de la technique de *poursuite de projection*, plus adaptée aux données sous-représentées. Néanmoins, cette technique est très coûteuse en temps de calcul.

D'autres versions robustes de l'ADL reposent sur le rééchantillonnage des données en entrée [FL03b, LJ03]. Les principales techniques de rééchantillonnage des données, ainsi que leur intérêt dans le cadre de l'ADL, sont présentées en annexe G (p. 205).

3.4.3.2 L'ADL résistante aux classes aberrantes

Les classes aberrantes [LDHU01] sont celles qui sont le plus *séparées* des autres classes, dans l'espace original des données. On considère qu'une classe est séparée des autres si sa moyenne est significativement différente des autres moyennes de classes, et que sa variance est telle qu'il n'existe pas ou peu de chevauchement avec les autres classes.

a) Robustifier le calcul de la variance inter-classe Décomposons la matrice de variance inter-classe à k classes en une somme de $\frac{1}{2}k(k-1)$ matrices de variance entre deux classes : on peut montrer [Loo00] que la matrice de variance inter-classe donnée en équation (3.5) peut

11. Sensibilité aux translations et/ou aux changements d'échelle

12. Le point d'effondrement d'un estimateur peut être vu comme la somme d'information aberrante (arbitrairement fixée à une valeur extrêmement grande) supportée par l'estimateur, sans que l'estimation fournie varie significativement. À titre d'exemple, il suffit d'assigner à une seule observation une valeur arbitrairement très grande, pour changer l'estimation de la moyenne de façon significative. Le point d'effondrement de la moyenne est donc nul, elle n'est pas résistante.

s'exprimer :

$$S_b = \frac{1}{N^2}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} N_j N_i(\overline{A_i}-\overline{A_j})(\overline{A_i}-\overline{A_j})^T \tag{3.17}$$

Ainsi réécrite, il est évident que cette matrice de variance est très largement influencée par les classes les plus distantes entre elles. La transformation résultant de l'ADL préserve donc la forte séparation des classes éloignées au détriment d'autres classes plus rapprochées, ce qui peut engendrer des chevauchements de ces dernières. Ces chevauchements sont source d'ambiguïté lors de la classification de nouveaux exemples, et peuvent engendrer une baisse des performances.

Pour corriger ce phénomène, il existe deux approches principales : la première, proposée par Loog *et al.* dans [LDHU01] et appelée ADL Pondérée (ADLP), consiste en une repondération dans le calcul de S_b avant d'appliquer l'ADL et la seconde, introduite par Lotlikar et Kothari dans [LK00], repose sur la décomposition de chaque étape de réduction de dimension en un certain nombre de pas fractionnaires (il s'agit de l'*approche par pas fractionnaires*).

L'ADL Pondérée (ADLP) Le but de cette technique est de construire une ADL basée sur une variante $\tilde{S_b}$ de la matrice de variance inter-classe S_b. La matrice $\tilde{S_b}$ est une version pondérée de la matrice S_b, la contribution de chaque classe étant proportionnelle à son taux de chevauchement avec ses voisines (erreur de Bayes entre classes). Sous l'hypothèse que la matrice de variance intra-classe S_w est la matrice unité, le critère à maximiser $J(W)$ adapté du critère de Fisher (3.9) est le suivant :

$$J(W) = \sum_{p=1}^{k-1} N_p \sum_{q=p+1}^{k} N_q \omega(\Delta_{pq}) \mathrm{trace}(W^T S_{pq} W) \tag{3.18}$$

où S_{pq} est la variance entre les centres des classes Ω_p et Ω_q :

$$S_{pq} = (\overline{A_p}-\overline{A_q})(\overline{A_p}-\overline{A_q})^T,$$

Δ_{pq} est la distance Euclidienne entre les moyennes de classes $\overline{A_p}$ et $\overline{A_q}$:

$$\Delta_{pq} = \sqrt{(\overline{A_p}-\overline{A_q})(\overline{A_p}-\overline{A_q})^T},$$

et le poids $\omega(\Delta_{pq})$ attribué à la paire de classes (Ω_p, Ω_q) est :

$$\omega(\Delta_{pq}) = \frac{1}{2\Delta_{pq}^2} F(\frac{\Delta_{pq}}{2\sqrt{2}}) \tag{3.19}$$

où la fonction

$$F(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)dt$$

est telle que $\frac{1}{2}(1+F(\frac{x}{\sqrt{2}}))$ est la fonction de répartition d'une loi $\mathcal{N}(0,1)$. Loog *et al.* montrent que le poids $\omega(\Delta_{pq})$ est une approximation de la précision moyenne du classifieur lors de la séparation des deux classes, chacune de ces classes étant distribuée selon une loi normale de variance unité. Le critère (3.18) est appelé *critère de Fisher paire à paire pondéré* [LDHU01].

La matrice W optimale au sens du critère (3.18) est composée des g premiers vecteurs propres (associés aux g plus grandes valeurs propres) de la matrice $\tilde{S_b}$, telle que :

$$\tilde{S_b} = \sum_{p=1}^{k-1} N_p \sum_{q=p+1}^{k} N_q \omega(\Delta_{pq}) S_{pq} \tag{3.20}$$

Loog *et al.* montrent que, si la matrice de variance intra-classe n'est pas la matrice identité, il suffit de blanchir les données, avant d'appliquer l'ADLP (calcul et analyse propre de $\tilde{S}_b$), puis de se ramener à l'espace original des données par projection. Cela revient à remplacer l'étape 2. de l'algorithme de résolution par diagonalisations successives, donné en section E.3.2 (p. 194) de l'annexe E, par la technique décrite ci-dessus.

Price et Gee [PG05] appliquent l'ADLP à la reconnaissance de visages. Pour surmonter le problème de la sous-représentativité des données, ils la mettent en œuvre de manière Directe [YY01] (voir section 3.3.3.3). Cet algorithme d'ADL Directe Pondérée sera désigné par l'acronyme ADLDP. Price et Gee introduisent également un algorithme d'ADL Pondérée précédée d'une phase d'ACP, désignée par l'acronyme ACP+ADLP.

L'approche par pas fractionnaires Considérons que l'on cherche à maximiser le critère suivant, équivalent au critère de Fisher (3.9) si la matrice S_w de covariance intra-classe est la matrice identité :

$$J(W) = \text{trace}(Sb) \tag{3.21}$$

Supposons que l'on veuille réduire la dimensionnalité du sous-espace discriminant de n à $n-1$. L'algorithme standard consiste à calculer les vecteurs propres W_i, $i = 1, \ldots, n$ de la matrice réelle symétrique S_b, puis à les ranger par ordre décroissant de leur valeur propre associée et enfin à projeter les exemples de la base d'apprentissage dans $\mathcal{F} = \text{vect}(W_1, W_2, \ldots, W_{n-1})$, le $n^{\text{ème}}$ vecteur propre W_n n'étant pas pris en compte. En présence d'un nombre important de classes, ou de classes aberrantes, il est possible qu'il existe deux classes, notées Ω_p et Ω_q, telles que la direction de la droite reliant leurs centres $(\overline{A_p} - \overline{A_q})$ soit approximativement la même que celle du vecteur W_n. Et, puisque $W_n \perp \mathcal{F}$, une fois que ces deux classes Ω_p et Ω_q sont projetées dans $\mathcal{F}$, leur chevauchement est très important.

Au lieu de simplement rejeter le $n^{\text{ème}}$ vecteur propre sans modifier les $n-1$ premiers, Lotlikar et Kothari proposent dans [LK00] de décomposer cette phase de rejet en un certain nombre de pas fractionnaires. À chaque pas, l'influence de la $n^{\text{ème}}$ composante W_n est réduite, et la matrice de transformation $W = [W_1, \ldots, W_n]$ est recalculée (par analyse propre de S_b). En d'autres termes, à chaque pas fractionnaire, le sous-espace se réoriente de manière à éviter les chevauchements entre classes. Au bout d'un certain nombre de pas fractionnaires, l'influence de la $n^{\text{ème}}$ composante devient nulle ; on peut la rejeter car l'information discriminante qu'elle contenait a été progressivement transférée vers le sous-espace de dimension $n-1$.

Afin de réduire l'influence de la $n^{\text{ème}}$ composante, on applique à chaque pas fractionnaire m une compression de la $n^{\text{ème}}$ coordonnée des observations projetées, selon :

$$\hat{A}_l(i) = \begin{cases} \alpha^m \hat{A}_l(i) & \text{si } i = n \\ \hat{A}_l(i) & \text{si } i \neq n \end{cases}$$

où $\hat{A}_l(i)$ est la $i^{\text{ème}}$ coordonnée du vecteur $\hat{A}_l$, projection de l'observation A_l sur le sous-espace discriminant $\text{vect}(W_1^m, \ldots, W_n^m)$ de l'étape m ; et la valeur de α^m, tel que $0 < \alpha^m < 1$, appelé *facteur de réduction*, décroît à chaque pas.

On remarquera que l'on peut appliquer cette technique à des cas où la matrice de variance intra-classe n'est pas la matrice identité, en utilisant une phase préliminaire de blanchiment des données, comme pour la technique pondérée de Loog *et al.* présentée ci-avant.

Cette technique connaît un certain succès dans le cadre de la reconnaissance de visages. Lu *et al.* [LPV03b] en ont introduit une version Directe (au sens de Yu et Yang [YY01]) (cette version

sera appelée ADL Directe Fractionnaire, son acronyme est ADLDF), tandis Dai *et al.* [DQJ04] en ont proposé une version à noyau (les techniques à noyau seront présentées en section 3.5). Néanmoins, le coût de calcul de la décomposition en pas fractionnaires est généralement très important.

b) Robustifier le calcul de la variance intra-classe Très récemment, Tang *et al.* [TSYQ05] ont proposé une technique permettant de diminuer l'influence des classes aberrantes dans le calcul de la matrice de variance intra-classe S_w. La matrice de variance intra-classe S_w est la moyenne pondérée des matrices V_j, pour $j = 1, \ldots, k$, où V_j est la matrice de variance de la classe Ω_j :

$$S_w = \frac{1}{N}\sum_{j=1}^{k} N_j V_j \quad \text{où} \quad V_j = \frac{1}{N_j}\sum_{A_l \in \Omega_j} (A_l - \overline{A_j})(A_l - \overline{A_j})^T, \text{ pour tout } j = 1, \ldots k$$

Considérons le cas extrême où des éléments issus de l'une de ces matrices V_j sont beaucoup plus importants que les éléments correspondants provenant des autres V_j. La matrice contenant les plus grands éléments aura alors une influence dominante sur la phase de blanchiment de S_w. Si la classe dominante Ω_j est simultanément une classe aberrante, le critère de Fisher usuel se focalise sur la minimisation de la variance intra-classe de celle-ci au détriment des autres classes, alors même qu'elle est aisément séparable des autres, puisqu'elle en est très éloignée.

Les classes les plus isolées doivent donc avoir moins d'influence dans le calcul de S_w. Par conséquent, la matrice de variance intra-classe S_w est transformée selon :

$$\tilde{S_w} = \frac{1}{N}\sum_{j=1}^{k} N_j \omega_j V_j$$

où le poids ω_j assigné à la classe Ω_j est inversement proportionnel à la distance moyenne entre Ω_j et les autres classes. Plusieurs mesures de dissimilarité, allant d'une simple distance Euclidienne au critère de Chernoff [LD04], sont envisagées. La distance utilisée tient compte de toutes ces distances, de manière pondérée. Un algorithme d'optimisation basé sur l'évolution (proche d'un algorithme génétique) est introduit dans [TSYQ05] pour sélectionner le vecteur de poids ω. Pour cela, une base de validation distincte de la base d'apprentissage est utilisée. L'utilisation de cet algorithme d'optimisation nous évite d'avoir à choisir une unique mesure de dissimilarité, qui pourrait ne pas être optimale dans tous les cas de figure.

3.4.3.3 Évaluation des performances des techniques proposées

Le tableau 3.4 donne une comparaison des performances des principales techniques présentées ci-dessus. Les résultats obtenus sur différentes bases de visages internationales montrent la supériorité des techniques d'ADL robuste, en particulier de l'ADL Directe Fractionnaire (ADLDF) et de l'ACP+ADL Pondérée (ACP+ADLP) sur les techniques non robustes.

3.5 L'Analyse Discriminante Linéaire à noyau

Dans le cas où les données sont trop complexes pour qu'une combinaison linéaire des caractéristiques permette de séparer les classes, une ADL peut ne pas suffire à définir des règles de classification satisfaisantes. Diverses généralisations non linéaires de l'ADL ont donc été introduites, parmi lesquelles les méthodes dites d'« ADL à noyau ». L'utilisation d'une fonction

Source	Base	Conclusion
[DQJ04]	ORL	ADLDF>EFM>ADLD>*fisherfaces*>*eigenfaces*
[LPV03b]	ORL	ADLDF>ADLD>*eigenfaces*>*fisherfaces*
	UMIST	ADLDF>ADLD>*fisherfaces*>*eigenfaces*
[PG05]	Yale+AR+ FERET+CVL	ACP+ADLP$\simeq$*fisherfaces*>ADLDP>ADLD>*eigenfaces*

TAB. 3.4 – *Comparaison des performances des algorithmes proposés, sur diverses bases de visages. Le symbole '$\simeq$' signifie que les performances des deux techniques sont comparables. L'opérateurs '$\geq$' désigne les cas où la première méthode est meilleure que la seconde, mais de manière non significative ; on utilisera le symbole '>' si la différence de performance est significative. Dans [PG05], la base utilisée pour l'évaluation est issue de l'agglomération de plusieurs bases.*

de noyau sert à ramener le problème de classification dans un espace $\mathcal{K}$ de très grande dimension d, où l'on espère qu'une ADL suffira à classer les données. Dans cette optique, Aizerman *et al.* [ABR64] ont baptisé l'espace $\mathcal{K}$ *espace de linéarisation*. On construit le classifieur depuis les données projetées dans l'espace $\mathcal{K}$; les facteurs discriminants ainsi obtenus sont des séparateurs linéaires dans l'espace $\mathcal{K}$ de très grande dimension, mais non linéaires dans l'espace initial des données. C'est pourquoi l'utilisation de la fonction de noyau rend dans une certaine mesure l'ADL non linéaire. La figure 3.6 illustre l'objectif de l'utilisation d'une fonction de noyau : en deux dimensions il est impossible de trouver une droite séparant les deux classes, tandis que si l'on considère la fonction :

$$\begin{array}{rccl} \Phi : & \mathbb{R}^2 & \rightarrow & \mathbb{R}^3 \\ & (x_1,x_2) & \mapsto & (z_1,z_2,z_3) := (x_1^2,\sqrt{2}\,x_1x_2,x_2^2) \end{array}$$

les données projetées dans la base transformée (z_1,z_2,z_3) sont telles que les deux groupes sont linéairement séparables. La règle de séparation, linéaire dans $\mathbb{R}^3$, correspond à une règle de décision non linéaire (borne ellipsoïdale) dans $\mathbb{R}^2$.

Au vu des difficultés engendrées par la sous-représentativité des données, on peut se demander s'il est pertinent de se ramener dans un espace $\mathcal{K}$ de très grande dimension d par rapport à la dimensionnalité n des données originales. En effet, cela revient à augmenter la dimensionnalité des données, pour un nombre d'observations N fixé, et donc à faire décroître le ratio entre le nombre d'observations et la taille de celles-ci. Le problème de sous-représentativité des données, déjà aigu dans le contexte de la reconnaissance de visages, se trouve encore aggravé. Néanmoins, la théorie de l'apprentissage statistique montre qu'il peut être plus aisé de construire un classifieur précis dans $\mathcal{K}$ que dans l'espace original des données, si la complexité du classifieur est moins importante [MMR+01]. Reprenons l'exemple donné en figure 3.6. Si nous avions voulu résoudre le problème dans $\mathbb{R}^2$, il nous aurait fallu construire un classifieur non linéaire complexe, alors que dans $\mathbb{R}^3$ une simple règle linéaire suffit à classer les données (et la différence de dimensionnalité entre $\mathbb{R}^2$ et $\mathbb{R}^3$ n'est pas drastique). Il existe donc deux paramètres à contrôler lors de l'utilisation d'une fonction de noyau : la complexité du classifieur choisi, et les dimensions du problème. L'espace $\mathcal{K}$ peut être de très grande taille ; dans ce cas, il sera très difficile de travailler directement

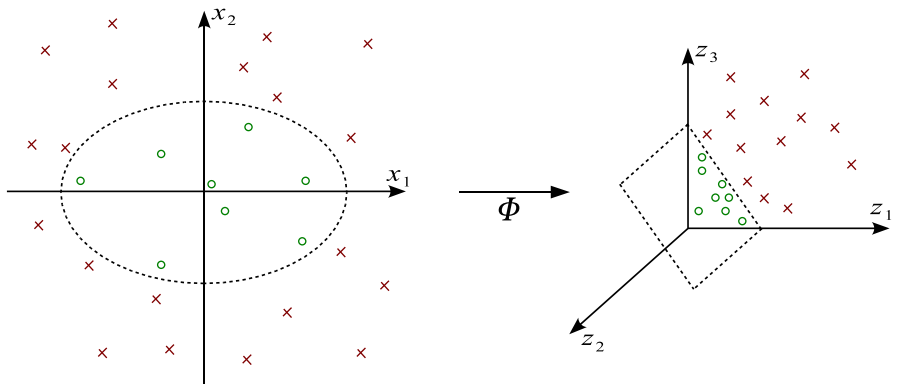


FIG. 3.6 – *Extrait de [SS01]. Les deux groupes (croix et ronds) ne sont pas linéairement séparables dans l'espace initial des données. Par contre, si l'on choisit une transformation* $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ *astucieuse, les données projetées* $\Phi(X)$ *sont linéairement séparables dans* $\mathbb{R}^3$*, ce qui correspond à une règle de décision non linéaire (borne ellipsoïdale) dans l'espace initial des données.*

dans cet espace. Heureusement, nous montrerons que, sous certaines conditions, il n'est pas nécessaire d'expliciter l'espace $\mathcal{K}$, ni la fonction Φ : seule l'utilisation d'un produit scalaire dans $\mathcal{K}$ est nécessaire. Ce produit scalaire K est tel que :

$$\begin{array}{rccl} K & : & \mathbb{R}^n \times \mathbb{R}^n & \rightarrow & \mathbb{R} \\ & & (X,Y) & \mapsto & K(X,Y) \;=\; \Phi(X) \cdot \Phi(Y) \;=\; \Phi(X)^T \Phi(Y) \end{array} \tag{3.22}$$

où $\cdot$ est le produit scalaire Euclidien de $\mathbb{R}^d$. Le produit scalaire K est appelé *fonction de noyau*.

3.5.1 Les fonctions de noyau

Si la fonction de noyau K vérifie les conditions du théorème de Mercer [Vap98], alors il existe un espace $\mathcal{K}$ et une application Φ de l'espace initial des données dans $\mathcal{K}$, tel que :

$$\begin{array}{rccc} \Phi & : & \mathbb{R}^n & \rightarrow & \mathcal{K} \\ & & X & \mapsto & \Phi(X) \end{array} \tag{3.23}$$

où $\mathcal{K}$ est de taille quelconque (sa taille peut même être infinie) et est tel que la fonction de noyau K implémente le produit scalaire $\cdot$ dans $\mathcal{K}$ (selon l'équation (3.22)). Nous montrerons par la suite que, sous les conditions du théorème de Mercer, il est inutile de déterminer la transformation Φ et l'espace $\mathcal{K}$: l'ADL à noyau peut être construite à l'aide de la fonction de noyau K uniquement.

Il existe de nombreuses fonctions de noyau vérifiant les conditions de Mercer [MMR+01]. Pour la reconnaissance de visages, les plus utilisés sont [SBV95] sont :

– le noyau gaussien :

$$K(X,Y) \;=\; \exp\left(-\frac{\|X-Y\|_2^2}{\sigma^2}\right) \tag{3.24}$$

où $\|\cdot\|_2$ est la norme Euclidienne et le paramètre $\sigma \in \mathbb{R}$;

– et le noyau polynomial :

$$K(X,Y) \;=\; (X \cdot Y + \theta)^d \tag{3.25}$$

où $\cdot$ est le produit scalaire Euclidien et les paramètres $(d,\theta) \in (\mathbb{N},\mathbb{R})$.

3.5.2 Description de la technique d'ADL à Noyau

3.5.2.1 Construction du modèle

Notons $A_l^\Phi = \Phi(A_l)$ les données de la base d'apprentissage projetées dans $\mathcal{K}$, $l = 1, \ldots, N$. On applique une ADL sur ces données. Le critère de Fisher devient :

$$J_\Phi(W) = \frac{|W^T S_b^\Phi W|}{|W^T S_w^\Phi W|} \tag{3.26}$$

où W est la matrice des facteurs discriminants, et les matrices S_w^Φ et S_b^Φ sont respectivement les matrices de covariance intra- et inter-classe des données projetées. On peut montrer que tout vecteur W_i de $\mathbb{R}^d$ (où d est la dimensionnalité de $\mathcal{K}$) maximisant le critère (3.26) repose dans le sous-espace expliqué par l'ensemble des données, projetées dans $\mathcal{K}$ [MRW+99]. Par conséquent, il existe un ensemble de coefficients $\alpha_1^{(i)}, \alpha_2^{(i)}, \ldots, \alpha_N^{(i)}$ tels que :

$$W_i = \sum_{l=1}^{N} \alpha_l^{(i)} \Phi(A_l) \tag{3.27}$$

Dans ces conditions, on peut montrer que le critère de Fisher donné en équation (3.26) devient :

$$J(\alpha) = \frac{|\alpha^T K_b \alpha|}{|\alpha^T K_w \alpha|} \tag{3.28}$$

où les vecteurs $\alpha^{(i)}$ (colonnes de la matrice α) maximisant ce critère sont les vecteurs propres de la matrice $K_w^{-1} K_b$ associés aux plus grandes valeurs propres, avec :

$$K_w = = \frac{1}{N} \sum_{j=1}^{k} \sum_{A_l \in \Omega_j} (Y_l - M_j)(Y_l - M_j)^T \tag{3.29}$$

où

$$Y_l = [K(A_1,A_l), K(A_2,A_l), \ldots, K(A_N,A_l)]^T \tag{3.30}$$

$$M_j = \left[\frac{1}{N_j} \sum_{A_m \in \Omega_j} K(A_1,A_m), \frac{1}{N_j} \sum_{A_m \in \Omega_j} K(A_2,A_m), \ldots, \frac{1}{N_j} \sum_{A_m \in \Omega_j} K(A_N,A_m) \right]^T \tag{3.31}$$

et

$$K_b = \frac{1}{N} \sum_{j=1}^{k} N_j (M_j - \overline{M})(M_j - \overline{M})^T \tag{3.32}$$

avec $\overline{M} = \frac{1}{N} \sum_{j=1}^{k} N_j M_j$ est la moyenne des M_j.

Lorsqu'un visage-requête X doit être classé, on commence par le projeter sur W :

$$W^T \Phi(X) = \sum_{l=1}^{N} \alpha_l K(A_l,X) \tag{3.33}$$

et les règles de classification les plus couramment retenues sont la règle du plus proche voisin et de la plus proche moyenne, selon la distance Euclidienne.

Notons que la matrice de variance K_w, de taille $N \times N$, est construite depuis au plus $N - k$ observations linéairement indépendantes. Par conséquent, son rang est au plus $N - k$, et elle n'est pas inversible. L'ADL ne peut donc être appliquée directement. Nous présenterons ci-après les principales solutions proposées et utilisées dans le contexte de la reconnaissance de visages.

Yang [Yan02] propose d'utiliser une phase intermédiaire d'ACP dans l'espace de linéarisation $\mathcal{K}$, avant d'appliquer l'ADL (ce qui revient à appliquer la technique des *fisherfaces* dans $\mathcal{K}$). Cette technique sera appelée par la suite *Fisherfaces* à Noyau (FishN).

Mika *et al.* [MRW+99] régularisent la matrice K_w par l'ajout d'un terme de régularisation diagonal σI_N. D'autres techniques de régularisation ont été présentées plus récemment [MSGA03, CYHD05], la dernière étant désignée par le terme d'ADL à Noyau Régularisée (ADLNR).

Liu *et al.* [LWLT04] introduisent une technique d'ADL dans l'espace nul de K_w, très proche de l'ADL_0 de Chen *et al.* [CLK+00] (*cf.* section 3.3.3.3) sur S_w. Cette méthode est désignée par l'acronyme ADLN_0.

Lu *et al.* [LPV03a] utilisent l'algorithme d'ADL Directe de Yu et Yang [YY01] dans l'espace de linéarisation $\mathcal{K}$. Cette technique sera appelée par la suite ADL du Noyau Directe (ADLND).

3.5.2.2 Choix de la fonction de noyau

Le choix de la fonction de noyau la plus adaptée à la classification de visages est un problème difficile. Gupta *et al.* [GAP+02] ont mené des expériences sur la base de visages UMIST (*cf.* annexe A) et ont trouvé que l'utilisation d'un noyau polynomial conduit à une baisse des performances comparé à une simple ADL. Ils ont également rapporté que les performances sont dégradées lorsque le degré du polynôme augmente. Par conséquent, ils préconisent d'utiliser un noyau Gaussien, mais montrent que les performances sont très dépendantes du paramètre σ. Néanmoins, dans d'autres travaux [LHLM02], on préfère l'utilisation d'un noyau polynomial. Il semblerait que le noyau le plus adapté varie en fonction de la base : dans [Yan02], l'utilisation de noyaux polynomiaux ou Gaussiens est sans effet pour la base ORL, tandis que pour la base de Yale le noyau Gaussien est plus performant. Dans [LLM04], Liu *et al.* introduisent un nouveau noyau, basé sur le noyau polynomial k, qu'ils appellent *noyau cosinus* :

$$K(X,Y) \quad = \quad \frac{k(X,Y)}{\sqrt{k(X,X)k(Y,Y)}}$$

et rapportent de bonnes performances.

3.5.3 Évaluation et comparaison des performances

La table 3.5 donne une comparaison des performances des algorithmes proposés pour l'ADL à noyau, et d'autres algorithmes tels que les *eigenfaces* et l'Analyse en Composantes Principales à Noyau (ACPN, voir section 2.2.2.1), l'Analyse en Composantes Indépendantes (ACI, voir section 2.2.2.5) et que les Machines à Vecteurs de Support (SVM, voir section 2.2.5). Pour la classification, généralement, c'est la distance Euclidienne au plus proche voisin qui est utilisée. Liu *et al.* [LLM04] montrent en effet que cette règle de classification engendre des résultats significativement meilleurs que la règle à la plus proche moyenne.

Source	Base	Conclusion
[Yan02]	ORL	FishN≥*fisherfaces* >ACPN≥*eigenfaces* >SVM≫ACI
	Yale A	FishN>*fisherfaces*≫SVM>ACPN>*eigenfaces*>ACI
[GAP$^+$02]	UMIST	FishN≦*fisherfaces*>ACPN≥*eigenfaces*
[CYHD05]	FERET	ADLNR≥ADLD≥ADLND≫*fisherfaces*>*eigenfaces*
	Yale B	ADLNR>ADLND>ADLD>*fisherfaces*>*eigenfaces*
	PIE	ADLNR>ADLND>ADLD>*fisherfaces*
[LWLT04]	ORL	ADL_0 ≥$ADLN_0$ >ADLD>*fisherfaces*
	FERET	ADL_0 ≦$ADLN_0$ >ADLD>*fisherfaces*
[LPV03a]	UMIST	ADLND>ACPN

TAB. 3.5 – *Comparaison des performances des algorithmes proposés, sur diverses bases de visages. Le symbole '≦' signifie que, suivant les données considérées, l'une des méthodes peut être meilleure que l'autre ou inversement. Le symbole '≥' désigne un cas où la première méthode est meilleure que la seconde, mais de manière non significative. Les opérateurs '>' et '≫' désignent respectivement les cas où la première méthode est significativement plus performante, et beaucoup plus performante, que la seconde. Cette distinction peut être subjective. Pour plus de détails, se référer aux articles sources.*

Cette table de comparaison inspire plusieurs remarques :

- Les techniques basées sur l'ADL (à noyau ou non) sont plus performantes que les techniques basées sur l'ACP (*eigenfaces* ou ACP à Noyau) [Yan02, GAP$^+$02, CYHD05, LPV03a], ainsi que l'Analyse en Composantes Indépendantes et les Machines à Vecteurs de Support [Yan02] ;
- La technique à noyau régularisée ADLNR semble plus performante que l'ADL dans le Noyau Directe (ADLND) [CYHD05]. La méthode d'$ADLN_0$ étant également plus performante que l'ADLND, on peut se demander laquelle de l'ADLNR ou de l'$ADLN_0$ est la technique le plus performante. Ces deux techniques n'ont jamais été comparées directement ; néanmoins, toutes deux ont été évaluées sur des sous-échantillons de la base de FERET. Avec $N_j = 2$ images par classe et $N = 250$ personnes pour l'ADLNR et $N = 70$ personnes pour l'$ADLN_0$, les taux de reconnaissance de l'ADLNR et de l'$ADLN_0$ sont respectivement de 79,9% et de 72,2%. Il semblerait donc que l'ADLNR soit plus performante que l'$ADLN_0$ (bien que, pour tirer une conclusion définitive, il faudrait comparer les deux méthodes en utilisant les mêmes images).

La conclusion principale que l'on peut tirer des résultats expérimentaux de l'état de l'art est que l'utilisation d'une fonction de noyau n'amène pas toujours une amélioration notable des performances par rapport à la version linéaire de l'algorithme [GAP$^+$02, LWLT04, CYHD05].

Dans certains cas, l'utilisation d'une fonction de noyau peut même engendrer une baisse des taux de reconnaissance [LWLT04, CYHD05].

3.6 Conclusion

Dans le contexte de la reconnaissance de visages, la plupart des techniques usuelles d'ADL sont basées sur une représentation unidimensionnelle des images de visages (par le biais de vecteurs). Dans ce cas, il est nécessaire de mettre en œuvre l'une des techniques d'ADL conçues pour surmonter le problème de la singularité et présentées en section 3.3. Or, celles-ci sont en général plus coûteuses et/ou nécessitent l'ajout de paramètres supplémentaires difficiles à ajuster par rapport à une ADL standard. Nous avons de plus montré en section 3.3.3.5 que les performances de ces techniques sont variables suivant les tailles et les caractéristiques des bases utilisées pour l'apprentissage et le test. Dans ces conditions, il paraît nécessaire d'investiguer d'autres modes de représentation des visages.

Dans ce chapitre, nous avons également étudié les méthodes robustes proposées dans les cas où les données sont hétéroscédastiques (les variances intra-classe sont différentes) ou dans les cas où l'on est en présence de données (ou de classes) aberrantes dans la base d'apprentissage. Nous avons aussi présenté les principales techniques visant, par l'utilisant d'une fonction de noyau, à mettre en œuvre une classification non linéaire. Si les techniques robustes apportent généralement une amélioration des taux de reconnaissance, ce n'est pas forcément le cas de l'utilisation d'une fonction de noyau qui, de plus, repose sur le choix difficile de la fonction de noyau la plus adaptée au problème.

Chapitre 4

Une nouvelle technique Discriminante Bidimensionnelle Orientée en monde fermé

4.1 Introduction

Dans la plupart des méthodes globales de reconnaissance de visages basées sur la projection statistique (voir section 2.2.2), les images de visages sont modélisées par le biais de vecteurs, avant d'appliquer une ou plusieurs techniques d'analyse de données, telles que l'Analyse en Composantes Principales (ACP) et/ou l'Analyse Discriminante Linéaire (ADL). Les images de visages, initialement sous la forme de matrices bidimensionnelles contenant leurs valeurs de pixels (en niveaux de gris), sont transformées en vecteurs par simple concaténation des lignes ou des colonnes de pixels. Cette modélisation unidimensionnelle (1D) engendre dans une certaine mesure la perte d'une partie de la structure bidimensionnelle des images initiales. De plus, la dimension des vecteurs-images ainsi obtenus est généralement très grande, ce qui pose un certain nombre de problèmes. En premier lieu, les matrices de covariance sont difficiles à estimer de manière précise à cause du faible nombre d'exemples dont on dispose, comparativement à la taille de ces exemples (problème de sous-représentation des données détaillé en section 3.3.2). Deuxièmement, plus les dimensions sont importantes, plus le coût de construction du modèle est élevé. En troisième lieu, le classifieur construit depuis un nombre insuffisant de données (au regard de leur dimension) peut être instable, au sens où de petits changements dans la base d'apprentissage peuvent engendrer des changements importants dans le modèle. Enfin, ce problème de sous-représentation des données empêche la mise en œuvre directe de l'Analyse Discriminante Linéaire, à cause notamment du problème de la singularité. Nous avons vu en section 3.3 que de nombreuses solutions ont été proposées pour ce problème. Néanmoins, la plupart d'entre elles sont coûteuses et/ou nécessitent l'ajout de paramètres difficiles à ajuster (par rapport à un modèle d'ADL standard) et leurs performances varient en fonction des caractéristiques des bases d'images utilisées.

En Janvier 2004, Yang *et al.* [YZFY04] ont introduit une technique qu'ils ont baptisée Analyse en Composantes Principales Bidimensionnelle (ACP2D), qui consiste à appliquer une ACP directement sur les images (matrices) de visages, utilisant pour cela une matrice de covariance généralisée calculée directement depuis les lignes des images de visages. La modélisation des données n'est donc pas totalement bidimensionnelle (comme pourrait le laisser penser le nom de la technique), mais bidimensionnelle orientée (2Do) en lignes (2DoL). Cette modélisation permet

de réduire le coût calculatoire et l'instabilité numérique lors de la construction du modèle. Nous montrerons de plus que la modélisation 2Do des données permet une tolérance accrue vis-à-vis de différentes sources de variabilité.

Or, nous avons montré au chapitre 2. (section 2.4) ainsi qu'au chapitre 3. (sections 3.4.3.3 et 3.5.3) que, dans le contexte d'une représentation 1D des visages, les techniques issues de l'ADL sont plus performantes que celles issues de l'ACP. Cela est imputable au fait que, tandis que l'ACP privilégie un critère de représentation des données, l'ADL, elle, cherche à maximiser une mesure de séparation entre classes et est donc plus adaptée à la classification. Afin d'allier le pouvoir discriminant de l'ADL aux avantages d'une représentation 2Do des données, nous introduisons dans ce chapitre une nouvelle technique d'extraction de caractéristiques appelée Analyse Discriminante Linéaire Bidimensionnelle orientée (ADL2Do). Celle-ci se décline en deux versions, selon que l'analyse statistique est menée sur les lignes ou les colonnes des images de visages. Ce mode de mise en œuvre de l'ADL est direct, au sens où il contourne implicitement le problème de la singularité, et ceci sans nécessiter la mise en œuvre de l'une des variantes coûteuses de l'ADL ni l'ajout de paramètres supplémentaires difficiles à ajuster. Après avoir sélectionné la mesure de dissimilarité la plus adaptée à la classification des signatures ainsi obtenues, nous montrerons l'efficacité de l'approche proposée dans le contexte de l'identification en monde fermé et la complémentarité des deux techniques issues de l'ADL2Do.

Ce chapitre est organisé comme suit. En section 4.2, nous présenterons les données dont on dispose et les principales notations utilisées dans le cadre de ce chapitre. En section 4.3, nous étudierons la technique d'ACP2D et nous montrerons la supériorité de la modélisation bidimensionnelle orientée (2Do) sur la représentation usuelle (1D) en termes de taux de reconnaissance, de coût de construction du modèle et de robustesse vis-à-vis de quelques-unes des principales sources de variation rencontrées dans le contexte de la reconnaissance de visages. Enfin, en section 4.4, nous introduirons et détaillerons les deux versions issues de l'Analyse Discriminante Linéaire 2D orientée (ADL2Do), mettrons en lumière ses très bonnes performances en comparaison avec les autres techniques de projection statistique, et montrerons leur complémentarité en termes de résultats de classification.

4.2 Données et notations

Pour construire le modèle d'extraction de signatures, nous disposons d'une base de connaissance Ω contenant N images de visages en niveaux de gris. Chaque image X_l est stockée sous la forme d'une matrice de pixels de taille $h \times w$. Les images sont centrées, de manière à ce que $\overline{X} = \frac{1}{N}\sum_{l=1}^{N} X_l = 0$. On considère que le nombre de personnes enregistrées (classes) est k. Chaque image de la base d'apprentissage est affectée à l'une de ces k classes $\Omega_1, \Omega_2, \ldots, \Omega_k$. Les matrices de covariance sont estimées depuis des *matrices de dispersion généralisées*, calculées directement à partir des images de visages. On dispose également d'une base de connaissance contenant les images de visages connus auxquels sont comparés les visages-requêtes. Chaque image de cette base est étiquetée par sa classe d'appartenance. La base de connaissance peut ou non être confondue avec la base d'apprentissage, suivant l'application considérée. Sauf indication contraire, cela sera le cas dans ce chapitre. Par contre, la base de connaissance est toujours distincte de la base de test, qui servira à l'évaluation de la capacité de généralisation des systèmes. Le tableau 4.1 résume les notations utilisées dans ce chapitre.

Notation	Description	Notation	Description
Ω	ensemble des images	Ω_j	classe j
h	nombre de lignes des matrices-images	w	nombre de colonnes des matrices-images
N	nombre d'images de Ω	k	nombre de classes
X_l	exemple (matrice-image) $l = 1, \ldots, N$	N_j	nombre d'images de Ω_j
$\overline{X}$	moyenne des exemples de Ω	$\overline{X_j}$	moyenne des exemples de Ω_j
S_T	matrice de covariance *totale généralisée* de Ω	g	nombre de vecteurs de projection du modèle
P	matrice de projection (4.1) $P \in \mathbb{R}^{w \times g}$	Q	matrice de projection (4.17) $Q \in \mathbb{R}^{h \times g}$
S_b	matrice de variance *inter-classe généralisée* (4.12)	S_w	matrice de variance *intra-classe généralisée* (4.14)
Σ_b	matrice de variance inter-classe généralisée (4.19)	Σ_w	matrice de variance intra-classe généralisée (4.20)

TAB. 4.1 – *Principales notations du chapitre 4.*

4.3 L'ACP Bidimensionnelle

4.3.1 Introduction

Généralement, l'ACP est appliquée sur les vecteurs-images (technique des *eigenfaces*). Les vecteurs-images de la base d'apprentissage étant de très grande dimension par rapport à leur nombre, le calcul de la matrice de covariance, donc de ses vecteurs propres, est instable. Bien que l'on puisse évaluer les vecteurs propres de la matrice de covariance sans passer par le calcul de celle-ci, en utilisant des techniques basées sur la Décomposition en Valeurs Singulières (SVD) [SK87, KS90], le problème d'imprécision n'est pas pour autant écarté puisque les vecteurs propres sont évalués statistiquement à partir de données souffrant de sous-représentation et ceci quelle que soit la méthode adoptée pour les estimer. De plus, la complexité en termes de coût de calcul pour déterminer ces vecteurs propres est très importante : le nombre d'opérations nécessaires est en $o([\min(hw,N)]^3)$, où hw est la très grande taille des vecteurs-images et N leur nombre (*cf.* section 2.2.2.1).

Dans le but de pallier ces inconvénients, Yang *et al.* ont introduit en Janvier 2004 l'Analyse en Composantes Principales Bidimensionnelle [YZFY04] qui, à la différence des méthodes usuelles de projection statistique, ne nécessite pas de transformation préalable des matrices-images en vecteurs-images. Une *matrice de covariance généralisée* est estimée directement depuis les matrices-images de la base d'apprentissage. L'analyse en éléments propres de cette matrice, qui est de taille très réduite par rapport à la matrice de covariance des *eigenfaces*, permet de déterminer les directions de projection de manière moins instable que les *eigenfaces*.

4.3.2 Description

Considérons une matrice de projection P de taille $w \times g$ (le paramètre g étant fixé) et la projection suivante :

$$X_l^P = X_l P \tag{4.1}$$

où X_l^P est la matrice de taille $h \times g$ correspondant à la projection de la matrice-image X_l sur P et constitue la signature associée au visage X_l par l'ACP2D. On cherche à déterminer la matrice P qui, pour une taille $w \times g$ donnée, maximise le critère $J(P)$ suivant mesurant la *dispersion généralisée* S_T^P des images de la base d'apprentissage projetées sur P selon (4.1) :

$$J(P) \quad = \quad \mathrm{tr}(S_T^P) \tag{4.2}$$

où, si l'on note $\overline{X^P} = \frac{1}{N}\sum_{X_l \in \Omega} X_l^P$ la moyenne des N matrices-images projetées, la matrice S_T^P s'écrit :

$$\begin{aligned} S_T^P &= \frac{1}{N}\sum_{X_l \in \Omega}\left[(X_l^P - \overline{X^P})^T (X_l^P - \overline{X^P})\right] \\ &= \frac{1}{N}\sum_{X_l \in \Omega}\left[(X_l P - \overline{X}P)^T (X_l P - \overline{X}P)\right] \\ &= P^T \left(\frac{1}{N}\sum_{X_l \in \Omega}\left[(X_l - \overline{X})^T (X_l - \overline{X})\right]\right) P \end{aligned} \tag{4.3}$$

Notons S_T la matrice de dispersion suivante, baptisée *matrice de covariance totale généralisée* :

$$S_T = \frac{1}{N}\sum_{X_l \in \Omega}\left[(X_l - \overline{X})^T (X_l - \overline{X})\right] \tag{4.4}$$

On peut montrer que la matrice S_T est définie positive. En utilisant la définition (4.4), le critère (4.2) peut se réécrire :

$$J(P) \quad = \quad \mathrm{tr}(P^T S_T P) \tag{4.5}$$

Ce critère est appelé *critère de dispersion totale généralisé*. On peut aisément montrer que les colonnes de la matrice $P = [P_1, \ldots, P_g]$ maximisant le critère (4.5) sont les vecteurs propres (orthonormés) de la matrice S_T, associés aux g plus grandes valeurs propres [YZFY04]. On considérera par la suite que les vecteurs propres P_i sont rangés dans P suivant l'ordre décroissant de leurs valeurs propres associées. Yang *et al.* ne proposent pas de méthodologie pour déterminer le nombre g optimal de vecteurs propres à retenir dans P.

La classification des visages passe par le calcul d'une distance entre leurs matrices-signatures et une règle d'affectation au plus proche voisin. La distance entre les deux signatures X_a^P et X_b^P des images X_a et X_b utilisée est la suivante et est appelée par la suite distance Euclidienne D_{L_2} :

$$D_{L_2}(X_a^P, X_b^P) = \sum_{i=1}^{g} d_{L_2}(X_a^{P_i}, X_b^{P_i}) \tag{4.6}$$

où d_{L_2} est la distance Euclidienne entre vecteurs (*cf.* équation (D.3) de l'annexe D) et $X_a^{P_i} = X_a P_i$ est la projection de l'image X_a sur le vecteur P_i, qui est le vecteur propre de la matrice S_T associé

à la $i^{\text{ème}}$ valeur propre. Supposons que l'on dispose d'une image-requête T, à laquelle on cherche à assigner une identité. Sa classification s'effectue en deux temps. On projette T sur P selon (4.1), de manière à obtenir sa signature T^P, puis on la compare à toutes les signatures de la base de connaissance avec une règle d'affectation au plus proche voisin : si

$$X_m \quad = \quad \underset{X_l \in \Omega}{\texttt{Argmin}} \left[D_{L_2}(T^P, X_l^P) \right]$$

et que X_m est associée à la classe Ω_j, alors on décide d'assigner à T l'identité Ω_j.

4.3.3 Évaluation des performances

Yang *et al.* ont mené dans [YZFY04] une évaluation poussée de l'ACP2D, notamment en utilisant les bases internationales AR et ORL (*cf.* annexe A). Dans toutes les expérimentations détaillées dans cette partie, les bases d'apprentissage servent également de bases de connaissance.

La base AR (*cf.* annexe A) est utilisée pour évaluer les performances de l'ACP2D en présence de variations dans les conditions d'illumination, dans le temps et dans les expressions faciales. 120 des 126 personnes enregistrées dans la base sont utilisées pour cette expérience. Pour chaque personne, on dispose de 14 vues collectées lors de deux sessions menées à quinze jours d'intervalle. Lors de chaque session, on considère 7 vues de chaque personne, dans des conditions variables d'illumination et d'expression faciale.

Dans le but d'évaluer l'effet de variations dans l'expression faciale, la base d'apprentissage est constituée de deux vues par personne (une par session), avec une expression faciale neutre. La base de test est constituée de six vues par personne, avec des variations dans l'expression faciale. Le taux de reconnaissance obtenu par ACP2D est de 96,1% (soit 692 bonnes classifications sur 720 requêtes), contre 94,7% (soit 682 bonnes classifications sur 720) pour les *eigenfaces*.

Afin d'évaluer l'effet de variations dans le temps, on utilise pour l'apprentissage les sept vues par personne acquises lors de la première session ; la base de test est constituée des sept vues de la seconde session. Le taux de reconnaissance atteint par l'ACP2D est de 67,6%, contre 66,2% pour les *eigenfaces*, soit une différence absolue entre leurs performances de 12 visages, à la faveur de l'ACP2D. Tout comme dans les évaluations menées par Gross *et al.* [GSC01] et détaillées en section 1.5.5, on note une baisse des taux de reconnaissance entre les deux sessions de la base AR. Néanmoins, le protocole expérimental retenu n'étant pas le même, on ne peut comparer directement les taux de reconnaissance obtenus par les deux auteurs.

Pour tester l'effet de changements d'illumination, on considère huit vues par personne : deux vues sous des conditions d'illumination neutres et six vues avec des variations dans les conditions d'illumination. On sélectionne au hasard parmi ces vues deux images par personne (une par session), qui constituent la base d'apprentissage. Les vues restantes composent la base de test. Cette opération est répétée seize fois. L'ACP2D permet d'obtenir en moyenne sur les seize partitions un taux de reconnaissance de 89,8%, bien meilleur que les 78% obtenus par les *eigenfaces*. En effet, cette différence représente en tout 85 visages sur les 720 testés. Il semble donc que l'ACP2D soit beaucoup plus performante que les *eigenfaces*, surtout en présence de variations dans les conditions d'illumination.

La base ORL est utilisée pour évaluer les performances de l'ACP2D en présence de variations limitées dans la pose de la tête, l'expression faciale et les conditions d'illumination ainsi que l'impact du nombre d'exemples par personne disponibles pour l'apprentissage. Les expériences sont

menées sur la totalité de la base, c.-à-d. dix vues par personne, pour les quarante personnes contenues dans la base. Les résultats expérimentaux montrent que l'ACP2D est significativement plus performante que les *eigenfaces*, l'ACP à noyau (*cf.* section 2.2.2.6) et l'ACI (*cf.* section 2.2.2.5). Yang *et al.* ont également comparé leur technique à celle des *fisherfaces* de Belhumeur *et al.* [BHK97]. Avec dix images par classe en moyenne pour l'apprentissage (stratégie d'évaluation de type *leave-one-out*[13]), les performances de l'ACP2D et des *fisherfaces* sont équivalentes avec respectivement 98,3% pour l'ACP2D et 98,5% pour les *fisherfaces*. Par contre, avec des nombres d'images par classe plus faibles, l'ACP2D est plus efficace que les *fisherfaces*.

4.3.4 Évaluation de la tolérance à différents facteurs de variabilité

Comme nous l'avons vu en section 1.5, il existe de multiples facteurs pouvant entraîner une baisse de performance des systèmes de reconnaissance. Dans la section précédente, nous avons montré que la modélisation 2Do des objets semble apporter une certaine tolérance à des variations dans le temps, l'expression faciale, la pose de la tête et les conditions d'illumination, par rapport aux techniques 1D usuelles (*eigenfaces*, *fisherfaces*, etc.). D'autres facteurs, aussi appelés *artefacts*, peuvent également mener à une baisse des taux de reconnaissance. Parmi ces artefacts on compte des imprécisions lors de la segmentation du visage, une mauvaise qualité des images et des occultations partielles. Dans cette section, nous présentons les résultats d'expérimentations visant à comparer la tolérance de l'ACP2D et des *eigenfaces* à ces artefacts. Ces expériences sont menées à l'aide de la base FERET (voir annexe A). Elles nous ont permis de déterminer, pour chacune des deux techniques évaluées, des Intervalles de Tolérance (IT) vis-à-vis de chacun des artefacts considérés. Cette étude a fait l'objet d'une publication dans [VGL04].

4.3.4.1 Protocole expérimental

Le protocole expérimental est le suivant : nous considérons des images extraites de FERET, montrant 200 personnes, avec une vue par personne. La plupart des visages arborent des expressions faciales neutres et aucune ne porte de lunettes. Pour chaque image, la position des yeux est connue (fournie avec la base) et est utilisée pour normaliser le visage. La normalisation, détaillée en annexe C, comporte quatre étapes :

1. rotation de l'image de manière à ce que les yeux soient alignés horizontalement ;
2. remise du visage à l'échelle, de manière à ramener la distance interoculaire à 70 pixels ;
3. découpage et redimensionnement de l'image à une taille de 150×130 pixels, le visage étant centré dans l'image ;
4. égalisation de l'histogramme de l'image.

Les 200 images normalisées forment la base d'apprentissage, aussi utilisée comme base de connaissance. L'impact de huit facteurs est étudié, chacun d'entre eux étant simulé par un paramètre (voir détails ci-après). Ces facteurs sont illustrés en figure 4.1. Les facteurs étudiés relèvent d'imprécisions lors de la phase de détection/segmentation des visages, d'une mauvaise qualité d'images ainsi que d'occultations partielles.

Les trois paramètres suivants simulent les effets d'imprécisions lors de la phase de détection/segmentation des visages :

- **Translations horizontales et verticales** : entre les étapes 2. et 3. de la normalisation, le visage est translaté dans l'image, soit horizontalement (de -30 à +30 pixels, c.-à-d. jusqu'à

13. À chaque étape, une image est retirée de la base d'apprentissage et est utilisée pour l'évaluation des performances. Cette opération est répétée N fois et le taux de reconnaissance moyen calculé sur ces N classifications.

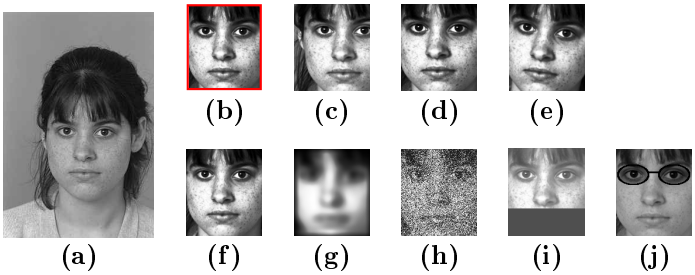


FIG. 4.1 – *(a) Image originale (tirée de FERET). (b) Image correctement normalisée. (c) Translation horizontale (+22 pixels). (d) Translation Verticale (+4 pixels). (e) Rotation (+8 degrés). (f) Changement d'échelle (-7%). (g) Lissage ($\sigma = 5{,}5$). (h) Ajout d'un bruit blanc gaussien ($\sigma = 90$). (i) Écharpe (47 pixels). (j) Lunettes ($\beta = 0{,}2$)*

23% de la largeur totale, les valeurs positives correspondant à une translation vers la droite), soit verticalement (de -19 à +19 pixels, c.-à-d. jusqu'à 12,7% de la hauteur totale, une valeur positive correspondant à une translation vers le haut). Les lignes ou les colonnes additionnelles (par rapport à la vue normalisée) sont issues de l'image originale ;

- **Rotation** : une rotation, de centre le milieu du segment interoculaire, est appliquée entre les étapes 2. et 3. de la normalisation. L'angle de rotation varie de 1 à 19 degrés, dans le sens des aiguilles d'une montre ;
- **Échelle** : durant l'étape 2., on fait varier la distance interoculaire de -20% à +20% par rapport à la valeur-étalon, c.-à-d. 70 pixels, ce qui engendre une variation de l'échelle du visage dans l'image.

Selon la distance entre le visage et l'appareil photographique, la résolution du visage à reconnaître peut être beaucoup plus faible que la résolution des visages de la base d'apprentissage. Dans ce cas de figure, une solution couramment adoptée est un zoom sur le visage, résultant en une interpolation qui peut engendrer un lissage de l'image. Le paramètre suivant simule ce phénomène :

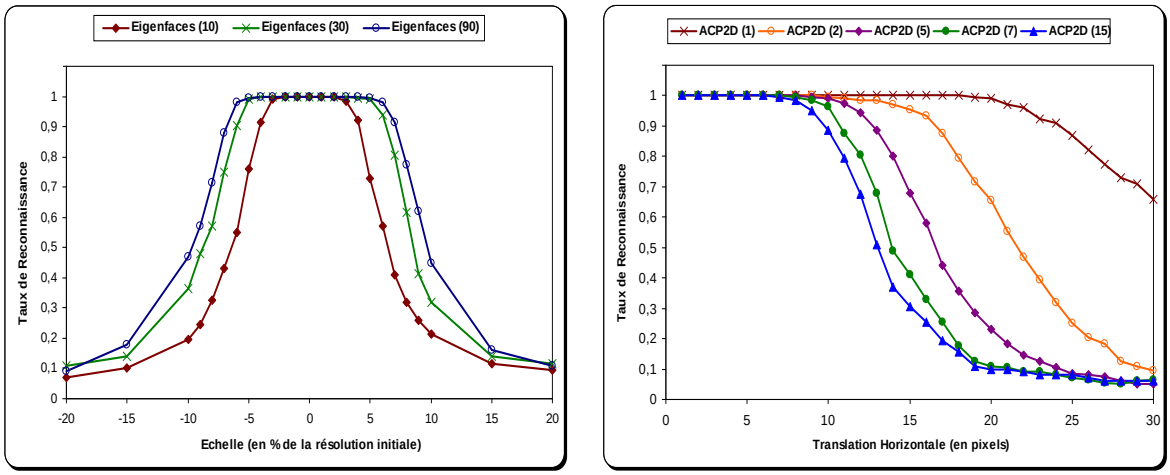
- **Lissage** : l'image est convoluée par un filtre gaussien, dont l'écart-type σ varie entre 0,5 et 9,5.

Les images acquises à l'aide d'appareils photographiques sont toujours contaminées par diverses sources de bruit ; nous faisons ici l'hypothèse que ce bruit est gaussien et nous le simulons par l'utilisation du paramètre suivant :

- **Bruit blanc** : un bruit blanc gaussien est ajouté à l'image entière, son écart-type varie de 1 à 90.

Nous étudions également les effets d'occultations partielles des visages, simulées par les paramètres suivants :

- **Écharpe** : une bande de pixels noirs, de largeur variant entre 1 et 80 pixels (jusqu'à 53% de la hauteur de l'image), couvre toute la largeur de l'image à partir du bas de l'image ;
- **Lunettes** : deux ellipses de trois pixels de largeur sont ajoutées à chaque image. Chacune est centrée sur un œil. Les longueurs des axes sont de 28 et 18 pixels. Les deux ellipses sont reliées par une bande noire de 3×17 pixels. Le niveau de gris $I(x,y)$ du pixel situé en (x,y) est remplacé par $I'(x,y) = (1-\beta)I(x,y)$. Le paramètre β varie de 0 à 1 : plus β augmente, plus l'intérieur des ellipses est foncé. Les lunettes sont donc complètement noires pour $\beta = 1$.



(a) **(b)**

FIG. 4.2 – *(a) Comportement de la technique des eigenfaces, en présence de variations d'échelle du visage dans l'image. (b) Comportement de l'ACP2D en fonction du nombre de vecteurs propres retenus, quand le visage est translaté horizontalement dans l'image. Les meilleurs taux de reconnaissance sont obtenus avec un seul vecteur propre. Les taux de reconnaissance décroissent rapidement lorsque l'on ajoute plus de vecteurs propres.*

Notre but est d'étudier les effets de chaque facteur de manière indépendante. Aussi, pour chaque expérimentation, on fait varier un seul paramètre. Chacune des valeurs prises par un paramètre définit une base de test, qui est comparée à la base d'apprentissage pour obtenir un taux de reconnaissance.

4.3.4.2 Résultats expérimentaux

Nous avons préalablement déterminé la valeur du nombre g de vecteurs propres permettant d'obtenir les meilleurs taux de reconnaissance pour la tâche considérée. Si l'on observe les taux de reconnaissance en fonction du nombre de vecteurs propres g retenus, deux cas de figure sont possibles : soit les taux de reconnaissance augmentent systématiquement avec le nombre de vecteurs propres (jusqu'à atteindre les limites fixées respectivement à 90 *eigenfaces* et à 20 vecteurs propres pour l'ACP2D), soit les taux de reconnaissance connaissent un pic autour d'une valeur g inférieure, que l'on peut considérer comme optimale. Le premier cas de figure, très largement observé pour les *eigenfaces*, est illustré en figure 4.2-a. Concernant l'ACP2D au contraire, comme l'illustre la figure 4.2-b, c'est le deuxième cas qui est plus largement observé. Ce point sera discuté en section 4.3.5. Dans les graphes de la figure 4.2, le nombre de vecteurs propres utilisés est systématiquement donné entre parenthèses (p. ex. ACP2D (6) signifie que le modèle utilisé repose sur l'ACP2D avec six vecteurs propres).

Les huit graphes de la figure 4.3 montrent l'évolution des taux de reconnaissance lorsque chaque paramètre est varié indépendamment. L'étude de ces graphiques nous permet de déterminer les *Intervalles de Tolérance* (IT) pour chacun des huit paramètres. Nous définissons l'Intervalle de Tolérance d'un classifieur pour un paramètre donné comme étant l'intervalle de variation du paramètre à l'intérieur duquel les taux de reconnaissance obtenus sont supérieurs à 95%. La table 4.2 donne, pour chaque technique, les IT à chacun des artefacts étudiés.

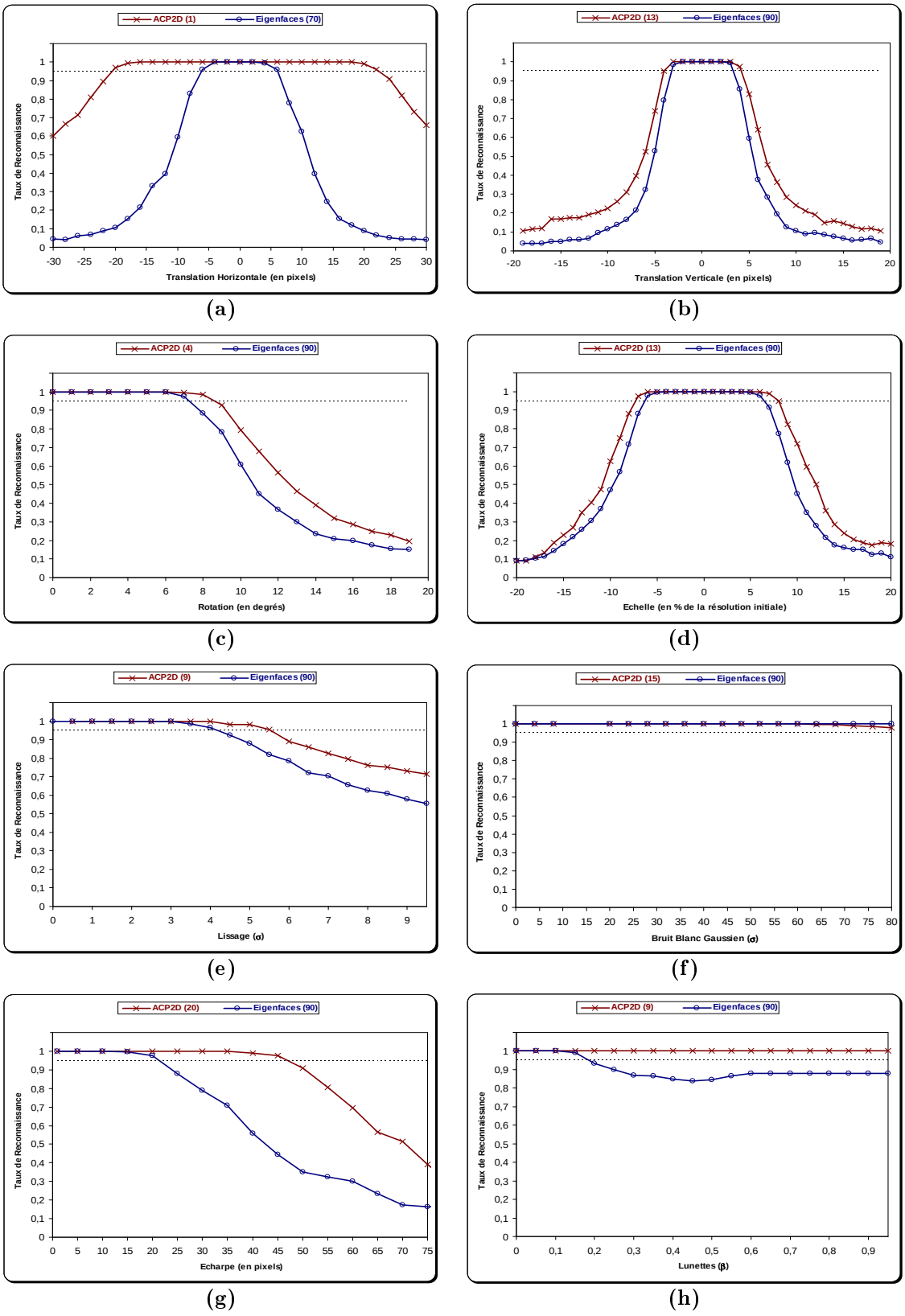


FIG. 4.3 – *Impact des huit paramètres étudiés sur les performances des eigenfaces et de l'ACP2D.*

	ACP2D	**Eigenfaces**
Translation Horizontale (en % de la larg. totale)	± 17%	± 4,6%
Translation Verticale (en % de la haut. totale)	± 2,7%	± 2%
Rotation (en degrés)	[0–8]	[0–6]
Échelle (en % de la résolution initiale)	± 7%	± 6%
Lissage (σ)	[0–5,5]	[0–4]
Bruit Blanc Gaussien (σ)	[0–90]	[0–90]
Écharpe (en % de la haut. totale)	31%	15%
Lunettes (β)	[0–1]	[0–0,15]

TAB. 4.2 – *Intervalles de Tolérance (IT) comparés de l'ACP2D et de la méthode des eigenfaces, pour les huit paramètres considérés.*

Concernant les translations horizontales (voir figure 4.3-a), l'ACP2D est beaucoup plus robuste que la méthode des *eigenfaces*, puisqu'elle permet d'obtenir un IT de [-20,22] pixels (environ 17% de la largeur totale de l'image) en utilisant uniquement le premier vecteur propre. Les *eigenfaces*, quant à elles, donnent un intervalle de tolérance de seulement [-6,6] pixels, soit moins de 4,6% de la largeur totale, et cela en utilisant 70 *eigenfaces*. Il est à noter que, lorsque l'on ajoute plus d'un vecteur propre, les taux de reconnaissance de l'ACP2D diminuent (voir figure 4.2-b), mais sans pour autant atteindre les faibles performances des *eigenfaces*. Une explication possible de ce phénomène est donnée en section 4.3.5.

La figure 4.3-b montre que l'ACP2D est bien moins tolérante aux translations verticales qu'aux translations horizontales. Elle est néanmoins plus tolérante que les *eigenfaces* : son IT est de ±2,7% de la hauteur totale, contre seulement ±2% pour les *eigenfaces*. Il faut également noter que le nombre de vecteurs propres nécessaires est beaucoup plus important que dans le cadre de translations horizontales. Cela pourrait être dû en partie à l'hypothèse formulée en section 4.3.5, à savoir que les premiers vecteurs propres de l'ACP2D encoderaient surtout de l'information concernant les positions verticales relatives des divers éléments faciaux. On peut également y voir les effets de la symétrie du visage. En effet, sous une vue frontale, il existe un axe de symétrie vertical passant par le nez. L'information provenant des lignes de l'image est donc redondante, ce qui explique que la plupart des systèmes de reconnaissance, à l'instar des *eigenfaces*, sont plus tolérants aux translations horizontales qu'aux translations verticales.

Les figure 4.3(c–e) montrent que l'ACP2D est également plus robuste que les *eigenfaces* aux rotations du visage dans l'image, à des changements d'échelle (positifs ou négatifs), ou encore au lissage de l'image (simulant la remise à l'échelle d'images de résolution initiale insuffisante).

La figure 4.3-f met en évidence le fait que les deux méthodes évaluées sont très tolérantes à l'ajout de bruit blanc Gaussien dans les images. Les taux de reconnaissance sont très proches de 100% pour toute valeur de σ allant de 0 à 90 et ceci bien que cette dernière valeur corresponde à un bruit très fort (voir figure 4.1-h). Cette constatation est imputable au fait que l'ACP, en analysant des données multinormales centrées, élimine en moyenne le bruit blanc.

Les figures 4.3(g–h) montrent que l'ACP2D est sensiblement plus robuste aux occultations partielles que la méthode des *eigenfaces*. En effet, tandis que les *eigenfaces* ne tolèrent que 22 pixels d'occultation de la région basse du visage, l'ACP2D est robuste jusqu'à 47 pixels, ce qui représente une amélioration de la tolérance d'environ 114%. Il est à noter que les performances observées pour les *eigenfaces* viennent confirmer les résultats expérimentaux de Gross *et al.* [GSC01] (voir section 1.5.4), selon lesquels les techniques statistiques globales 1D, telles que les *eigenfaces* et les *fisherfaces*, sont peu robustes à l'occultation de la région basse du visage. L'ACP2D, caractérisée par une tolérance accrue, ne souffre pas du même inconvénient. L'ACP2D corrige donc ce désavantage des techniques unidimensionnelles. De plus, concernant les lunettes, le taux de reconnaissance de l'ACP2D est de 100% pour β variant de 0 à 1. L'ACP2D est donc beaucoup plus tolérante à des occultations des yeux que la technique des *eigenfaces*, dont l'IT est seulement de $[0,0{,}15]$.

4.3.5 Discussion

Dans cette section, nous cherchons à mieux cerner le comportement de l'ACP2D et à souligner ses avantages comme ses inconvénients, en comparaison avec son pendant 1D, à savoir la technique des *eigenfaces*.

Revenons sur la tolérance de l'ACP2D aux translations horizontales. Avec un seul vecteur propre, l'intervalle de tolérance est étonnamment large ($\pm$17%). Plus on rajoute de vecteurs propres, plus la robustesse décroît. La visualisation de l'information contenue dans ces derniers peut nous aider à expliquer ce comportement. À la différence des *eigenfaces*, les vecteurs propres obtenus par ACP2D sont de longueur w et non hw. Ils ne peuvent donc pas être directement visualisés sous la forme d'images de même résolution que les images initiales, comme c'était le cas pour les *eigenfaces* (*cf.* figure 2.2). Nous pouvons cependant étudier les résultats de reconstruction obtenus grâce à l'ACP2D. En effet, comme toute méthode de projection orthonormale, l'ACP2D peut être utilisée avec succès pour la compression des images de visages [ZSL05]. L'image projetée X_l^P (de dimension réduite) et la matrice de projection P peuvent être combinées pour obtenir une reconstruction $\hat{X}_l$ de l'image initiale X_l, selon :

$$\hat{X}_l = X_l P P^T = X_l^P P^T \tag{4.7}$$

Des exemples de reconstruction, avec un nombre de vecteurs propres variant de 1 à 10, sont donnés en figure 4.4. On voit que, si le premier vecteur propre est essentiellement représentatif des positions verticales des éléments faciaux (yeux, nez, bouche), les vecteurs suivants encodent plus de détails (information de plus haute fréquence), y compris concernant les positions horizontales de ces éléments. La prise en compte de ces détails dans le modèle engendre, en présence d'une translation horizontale du visage dans l'image, une distance plus importante entre les projections du visage original et du visage translaté. La prise en compte de plus d'un vecteur propre conduit donc à une baisse des performances du modèle.

Remarquons qu'appliquer une ACP2D sur les images de visages revient à appliquer une ACP sur l'ensemble des lignes des visages [WWZF05]. L'ACP2D est en fait 2D-orientée en Lignes et peut être rebaptisée ACP2DoL. Il suffit d'appliquer l'ACP2D non pas sur les matrices-images originales, mais sur les transposées de celles-ci, pour obtenir une ACP sur les colonnes des images et ainsi définir une ACP2D orientée en Colonnes (ACP2DoC). Notons que, si l'ACP2DoC est plus robuste aux translations verticales que les *eigenfaces* et que l'ACP2DoL, elle n'y est pas aussi tolérante que l'ACP2DoL pour les translations horizontales. Cela est imputable à la symétrie du

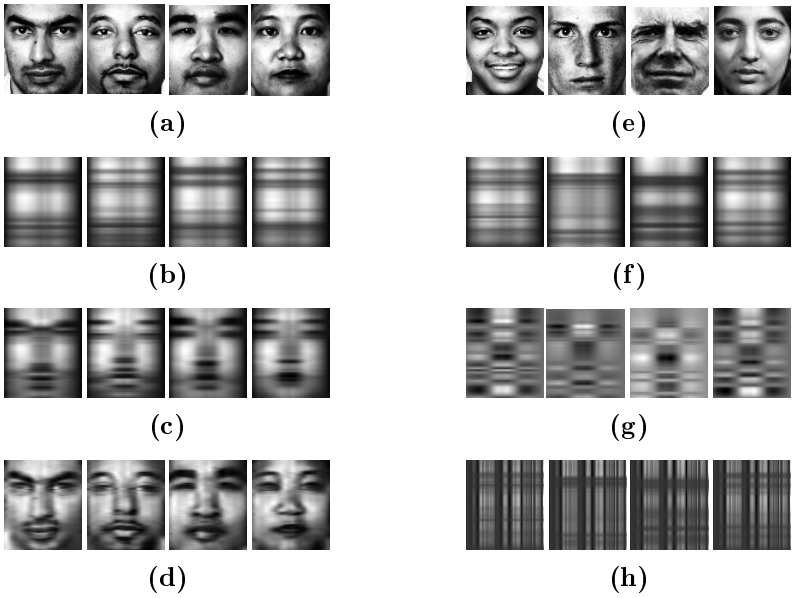


FIG. 4.4 – *(a,e) Images originales. La colonne de gauche montre les reconstructions obtenues avec les g premiers vecteurs propres et la colonne de droite montre les reconstructions obtenues avec le* $n^{\text{ème}}$ *vecteur propre uniquement, où (b,f) :* $g = 1$*, (c,g) :* $g = 2$ *et (d,h) :* $g = 10$*.*

visage déjà évoquée plus haut.

Comparons maintenant l'ACP2D et la technique des *eigenfaces*. Un avantage considérable de l'ACP2D sur cette dernière est qu'elle est beaucoup moins coûteuse en termes de nombre de calculs dans la construction du modèle. En effet, la taille de la matrice de dispersion dont on recherche les vecteurs propres est très réduite : $w \times w$ pour l'ACP2D contre $N \times N$ pour les *eigenfaces* (ou $hw \times hw$ selon que l'on utilise ou non l'astuce détaillée en section 2.2.2.1). La table 4.3, extraite de [YZFY04], compare les temps CPU nécessaires à la construction des modèles d'ACP2D et des *eigenfaces* (sur la base ORL). On peut noter que le gain apporté par l'ACP2D est d'autant plus flagrant que le nombre d'exemples est important. La taille réduite de la matrice de dispersion présente également un autre avantage, à savoir que son calcul est plus stable numériquement.

Nombre d'exemples / classe	**1**	**2**	**3**	**4**	**5**
Eigenfaces	44,45	89	139,36	198,95	304,61
ACP2D	10,76	11,23	12,59	13,40	14,03

TAB. 4.3 – *extrait de [YZFY04] - Comparaison des temps CPU pour la construction des modèles d'ACP2D et des eigenfaces, en utilisant cinq images par personne pour la base ORL (CPU : Pentium II 800 MHz, RAM : 256 Mb)*

Néanmoins, l'ACP2D présente un inconvénient comparé aux *eigenfaces* : le nombre de coefficients des signatures issues de l'ACP2D est beaucoup plus important que pour les *eigenfaces*. A titre d'exemple, pour la base ORL (*cf.* annexe A) et cinq images par personne dans la base

d'apprentissage, $g = 8$ vecteurs propres sont nécessaires pour obtenir des performances optimales avec l'ACP2D, contre $g = 100$ pour les *eigenfaces* [YZFY04]. Puisque la taille des images issues d'ORL est de 112×92 pixels, chaque signature fournie par l'ACP2D est une matrice de coefficients de taille $112 \times 8 = 896$, contre un vecteur de taille 100 seulement pour les *eigenfaces*. La phase de classification, qui consiste à mettre en correspondance les signatures de visages à l'aide d'une distance au plus proche voisin, est donc plus coûteuse en termes de temps de calcul pour l'ACP2D que pour les *eigenfaces*. Ceci constitue un désavantage important de l'ACP2D, car l'étape de classification est la plupart du temps menée en ligne (*cf.* figure 1). Afin de pallier ce problème, Yang *et al.* proposent d'appliquer une phase postérieure d'ACP, ce qui peut engendrer la perte d'une partie de l'information discriminante contenue dans les signatures initiales.

4.3.6 Conclusion

Dans cette partie, nous avons montré que la représentation 2Do des données dote l'ACP2D d'un certain nombre d'avantages, en comparaison avec la méthode des *eigenfaces* :

1. l'instabilité numérique est moins importante pour l'ACP2D que pour les *eigenfaces*, en raison de la taille réduite de la matrice de dispersion considérée ;
2. le coût calculatoire pour la construction du modèle d'ACP2D est beaucoup moins important que pour les *eigenfaces* ;
3. l'ACP2D est plus efficace que les *eigenfaces* pour la reconnaissance des visages, comme le montrent leurs résultats comparés sur plusieurs bases internationales (voir section 4.3.3) ;
4. l'ACP2D est plus tolérante que les *eigenfaces*, vis-à-vis d'une localisation imprécise des visages, d'une mauvaise qualité d'image ou encore d'occultations partielles.

Néanmoins, Yang [YZFY04] ne propose pas de stratégie de sélection du nombre de vecteurs propres à retenir. De plus, la taille des signatures fournies est généralement beaucoup plus importante que pour les *eigenfaces*. Enfin, la technique d'analyse de données mise en œuvre, à savoir l'ACP, est conçue dans un but de représentation (compression) des données, mais pas pour leur classification. Nous avons mis en lumière aux chapitres 2. et 3. que l'Analyse Discriminante Linéaire (ADL) est plus adaptée que l'ACP à cette tâche.

4.4 L'Analyse Discriminante Linéaire Bidimensionnelle Orientée

4.4.1 Introduction

Dans la section précédente, nous avons mis en évidence le fait que, pour l'ACP, la représentation 2Do des données apporte de nombreux avantages par rapport à une modélisation 1D (représentation des visages par le biais de vecteurs). Cependant, si l'ACP minimise l'erreur Euclidienne de reconstruction (voir section 2.2.2.1), rien ne prouve son efficacité pour classer efficacement les visages en fonction de leur identité. Par contre, nous avons montré au chapitre 2. (section 2.4) ainsi qu'au chapitre 3. (sections 3.4.3.3 et 3.5.3) que l'Analyse Discriminante Linéaire (ADL) est plus efficace pour la classification des visages que l'ACP. Nous introduisons donc dans cette partie une méthode de projection statistique alliant les avantages de la représentation 2Do des visages et le pouvoir discriminant de l'Analyse Discriminante Linéaire (ADL) : il s'agit de la technique d'*Analyse Discriminante Bidimensionnelle orientée* (ADL2Do) [VGJ04].

Nous avons vu en section 3.3 que, si les visages sont représentés de manière 1D, on ne peut appliquer directement l'ADL sur ces données à cause du problème de la singularité. Diverses solutions ont été présentées. La méthode des *fisherfaces* consiste à appliquer une phase préliminaire

de réduction de dimension. Mais nous avons vu que celle-ci peut conduire à une perte d'information discriminante. Différentes techniques d'ADL sous-optimale, conduisant pour la plupart au rejet d'une partie de l'information discriminante, ont également été introduites. Il existe enfin des techniques basées sur une modification du critère de Fisher, qui reposent généralement sur des paramètres variant en fonction des bases utilisées et difficiles à ajuster. De plus, selon les caractéristiques des bases utilisées pour l'évaluation, les performances de la plupart de ces techniques varient fortement (*cf.* section 3.3.3.5). Nous montrerons dans cette section que l'ADL2Do, quant à elle, peut être appliquée directement sur les images de visages. En effet, elle permet de contourner le problème de la singularité, ce qui évite d'avoir à recourir à ces solutions coûteuses ou pouvant engendrer une perte d'information discriminante. De plus, l'ADL2Do ne nécessite l'ajout d'aucun paramètre par rapport à une technique d'ADL standard. Nous montrerons sa supériorité sur les principales techniques de projection statistique 1D, ainsi que sur l'ACP2D.

Cette section est organisée comme suit. En section 4.4.2, nous décrirons mathématiquement et interpréterons géométriquement les deux versions de l'ADL2Do. Nous mènerons en section 4.4.3 une étude pour déterminer la mesure de dissimilarité la mieux adaptée à la classification des signatures obtenues. Nous introduirons en section 4.4.4 les différents modes de sélection du nombre de vecteurs de projection à considérer. Les effets de différents facteurs, tels que la résolution des images ainsi que le nombre de classes et d'exemples par classe, sont étudiés en section 4.4.5. Nous montrerons en section 4.4.6 que la technique proposée est à la fois plus efficace que l'ACP2D et que les principales techniques 1D basées sur l'ADL. En section 4.4.7, nous mettrons en lumière la complémentarité en termes de résultat de classification des deux techniques issues de l'ADL2Do. En section 4.4.8, nous discuterons de leurs principaux avantages et inconvénients, en comparaison avec ceux des approches statistiques de l'état de l'art.

4.4.2 Extraction de signatures

L'ADL2Do est une technique globale d'extraction de signatures se déclinant en deux versions : l'ADL2D orientée en lignes (ADL2DoL) et l'ADL2D orientée en colonnes (ADL2DoC). Par la suite, ces deux méthodes seront regroupées sous l'appellation d'ADL2Do. On utilise les mêmes notations que dans la section précédente. À la différence de l'ACP2D, l'algorithme de l'ADL2Do est *supervisé*, c.-à-d. que l'on connaît l'identité du visage courant (issu de la base d'apprentissage). La base d'apprentissage contient k classes $(\Omega_j)_{j=\{1\ldots k\}}$. La classe (Ω_j) contient N_j vues d'un même visage, avec $\sum_{j=1}^{k} N_j = N$. Présentons dans un premier temps l'ADL2DoL.

4.4.2.1 L'ADL2DoL

Considérons la projection (4.1), à savoir : $X_l^P = X_l P$, où X_l est une observation issue de la base d'apprentissage Ω et X_l^P est la matrice de longueur h, résultat de la projection de X_l sur la matrice P, de taille $w \times g$. Nous recherchons la matrice P, maximisant par projection la séparation des classes différentes tout en minimisant les variations intra-classe, pour une taille g fixée. Sous un certain nombre d'hypothèses que nous détaillerons en section 4.4.2.4, on peut considérer que P maximise le critère de Fisher généralisé suivant :

$$J_l(P) \quad = \quad \frac{|S_b^P|}{|S_w^P|} \tag{4.8}$$

où $|\cdot|$ est le déterminant et S_w^P et S_b^P sont respectivement les matrices de *dispersion intra-classe généralisée* et de *dispersion inter-classe généralisée* des images projetées selon (4.1), estimées à

partir des images de $\Omega = \{\Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_k\}$:

$$S_w^P = \frac{1}{N}\sum_{j=1}^{k}\sum_{X_l \in \Omega_j}(X_l^P - \overline{X_j^P})^T(X_l^P - \overline{X_j^P}) \tag{4.9}$$

$$S_b^P = \frac{1}{N}\sum_{j=1}^{k} N_j(\overline{X_j^P} - \overline{X^P})^T(\overline{X_j^P} - \overline{X^P}) \tag{4.10}$$

où $\overline{X_j^P} = \frac{1}{N_j}\sum_{X_l \in \Omega_j} X_l^P$ est la matrice moyenne des N_j images projetées de la classe Ω_j et $\overline{X^P} = \frac{1}{N}\sum_{j=1}^{k} N_j\overline{X_j^P}$ est la moyenne de tous les visages projetés de Ω. En introduisant l'équation (4.1) dans l'expression (4.9), on obtient :

$$\begin{aligned} S_w^P &= \frac{1}{N}\sum_{j=1}^{k}\sum_{X_l \in \Omega_j}(X_lP - \overline{X_j}P)^T(X_lP - \overline{X_j}P) \\ &= \frac{1}{N}\sum_{j=1}^{k}\sum_{X_l \in \Omega_j} P^T(X_l - \overline{X_j})^T(X_l - \overline{X_j})P \\ &= P^T S_w P \end{aligned} \tag{4.11}$$

où S_w est l'estimation sur Ω de la matrice de *covariance intra-classe généralisée* des observations :

$$S_w = \frac{1}{N}\sum_{j=1}^{k}\sum_{X_l \in \Omega_j}(X_l - \overline{X_j})^T(X_l - \overline{X_j}) \tag{4.12}$$

Par un cheminement analogue, on obtient :

$$S_b^P = P^T S_b P \tag{4.13}$$

où S_b est l'estimation sur Ω de la matrice de *covariance inter-classe généralisée* des visages :

$$S_b = \frac{1}{N}\sum_{j=1}^{k} N_j(\overline{X_j} - \overline{X})^T(\overline{X_j} - \overline{X}) \tag{4.14}$$

Le critère (4.8) devient donc :

$$J_l(P) = \frac{|P^T S_b P|}{|P^T S_w P|} \tag{4.15}$$

Si $g = 1$, maximiser le critère (4.15) revient à rechercher le vecteur P_1 de longueur w qui maximise la forme quadratique $P^T S_b P$. On peut considérer sans perte de généralité que la contrainte $P^T S_w P = 1$ est vérifiée. En effet, soit P' le vecteur maximisant le rapport (4.15) et vérifiant $P'^T S_w P' = c \neq 1$. Comme nous le verrons en section 4.4.2.3, la matrice S_w est régulière, donc ses valeurs propres sont non nulles et par conséquent $c \neq 0$. Il suffit alors de poser $P_1 = \frac{1}{\sqrt{c}}P'$ pour obtenir le maximum de $P^T S_b P$, la contrainte étant respectée. La recherche du maximum implique l'annulation des dérivées du Lagrangien :

$$L = P^T S_b P - \lambda(P^T S_w P - 1)$$

D'où on déduit la relation :

$$S_b P = \lambda S_w P$$

qui peut se réécrire, à condition que la matrice S_w soit inversible :

$$S_w^{-1} S_b P = \lambda P \tag{4.16}$$

Le vecteur de projection le plus discriminant P_1 est donc le vecteur propre de la matrice $S_w^{-1}S_b$ associé à la plus grande valeur propre λ_1.

En pratique, dans un contexte de classification multi-classes de données multivariées, un seul axe discriminant ne suffit pas pour obtenir des performances optimales. Comme pour l'ADL standard, nous retenons donc les g vecteurs propres P_1,P_2,...,P_g de $S_w^{-1}S_b$, correspondant aux plus grandes valeurs propres. La matrice $S_w^{-1}S_b$ n'étant pas nécessairement symétrique, on ne cherche pas à résoudre directement son système propre, mais on utilise l'algorithme de Fukunaga, détaillé en p. 194 de l'annexe E. Celui-ci, mettant en œuvre les diagonalisations successives des matrices de covariance intra- et inter-classe généralisées, nous permet d'obtenir l'ensemble des vecteurs propres non nuls de la matrice $S_w^{-1}S_b$. On sélectionne les g premiers vecteurs propres, c.-à-d. ceux qui sont associés aux plus grandes valeurs propres. On range ces vecteurs dans la matrice P, de taille $w \times g$, par ordre décroissant. Le mode de sélection du nombre g de vecteurs propres à retenir sera détaillé en section 4.4.4. L'ADL2DoL, une fois construite, permet donc d'assigner à chaque image X_l de Ω une matrice X_l^P, de taille $h \times g$, qui constitue la signature associée à cette image par l'ADL2DoL. Ce sont les signatures des images qui serviront à leur classification (*cf.* section 4.4.3).

4.4.2.2 L'ADL2DoC

La projection considérée dans le cadre de l'ADL2DoC est la suivante :

$$X_l^Q = Q^T X_l, \tag{4.17}$$

où Q est une matrice de projection de taille $h \times g$, Q^T est sa transposée et la matrice X_l^Q, de taille $g \times w$, est la projection selon (4.17) de X_l sur Q. Le critère (à maximiser) mesurant le pouvoir discriminant de la matrice Q est le suivant :

$$J_c(Q) = \frac{|Q^T \Sigma_b Q|}{|Q^T \Sigma_w Q|} \tag{4.18}$$

où les matrices Σ_w et Σ_b sont les estimations, calculées à partir des matrices de covariance intra- et inter-classe généralisées des matrices-images transposées X_l^T :

$$\Sigma_w = \frac{1}{N} \sum_{j=1}^{k} \sum_{X_l \in \Omega_j} (X_l - \overline{X_j})(X_l - \overline{X_j})^T \tag{4.19}$$

$$\Sigma_b = \frac{1}{N} \sum_{j=1}^{k} N_j (\overline{X_j} - \overline{X})(\overline{X_j} - \overline{X})^T \tag{4.20}$$

Par un raisonnement analogue à celui de l'ADL2DoL, on retient les g vecteurs propres de $\Sigma_w^{-1}\Sigma_b$ associés aux plus grandes valeurs propres. Ceux-ci constituent les colonnes de la matrice Q. Ainsi, la signature assignée à une image X_l par l'ADL2DoC est une matrice X_l^Q de taille $g \times w$.

4.4.2.3 L'ADL2Do et le problème de la singularité

Dans cette section, nous montrerons comment l'ADL2Do permet d'éviter implicitement le problème de la singularité, en augmentant artificiellement le ratio nombre d'observation / dimensionnalité ($\frac{N}{n}$) par rapport à une technique d'ADL basée sur une modélisation 1D des données.

Commençons par montrer que les rangs R et R' des matrices S_w (*cf.* équation 4.12) et Σ_w (*cf.* équation 4.14) vérifient respectivement $R \leq \min(w, (N-k)h)$ et $R' \leq \min(h, (N-k)w)$. Considérons le cas de l'ADL2DoL. Notons :

$$A_l = (X_l - \overline{X_j}) = \begin{bmatrix} A_l[1,1] & \dots & A_l[1,w] \\ \vdots & \ddots & \vdots \\ A_l[h,1] & \dots & A_l[h,w] \end{bmatrix}$$

Si l'on note $A_l[r]$ le vecteur de longueur w correspondant à la $r^{\text{ème}}$ ligne de la matrice A_l, alors cette dernière peut se réécrire :

$$A_l = \begin{bmatrix} A_l[1]^T \\ \vdots \\ A_l[h]^T \end{bmatrix}$$

La matrice de dispersion intra-classe généralisée S_w, donnée en équation (4.12) devient donc :

$$\begin{aligned} S_w &= \frac{1}{N}\sum_{j=1}^{k}\sum_{X_l\in\Omega_j}(X_l-\overline{X_j})^T(X_l-\overline{X_j}) = \frac{1}{N}\sum_{j=1}^{k}\sum_{A_l\in\Omega_j}A_l^TA_l \\ S_w &= \frac{1}{N}\sum_{j=1}^{k}\sum_{A_l\in\Omega_j}\begin{bmatrix} \sum_{r=1}^{h}(A_l[r,1])^2 & \sum_{r=1}^{h}A_l[r,1]A_l[r,2] & \dots & \sum_{r=1}^{h}A_l[r,1]A_l[r,h] \\ \sum_{r=1}^{h}A_l[r,1]A_l[r,2] & \sum_{r=1}^{h}(A_l[r,2])^2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{r=1}^{h}A_l[r,1]A_l[r,w] & \sum_{r=1}^{h}A_l[r,2]A_l[r,w] & \dots & \sum_{r=1}^{h}(A_l[r,w])^2 \end{bmatrix} \\ S_w &= \frac{1}{N}\sum_{j=1}^{k}\sum_{A_l\in\Omega_j}\sum_{r=1}^{h}A_l[r]A_l[r]^T \end{aligned}$$

Ce qui peut être reformulé comme suit :

$$S_w = \frac{1}{N}\sum_{j=1}^{k}\sum_{X_l\in\Omega_j}\sum_{r=1}^{h}(X_l[r]-\overline{X_j}[r])(X_l[r]-\overline{X_j}[r])^T \tag{4.21}$$

où $X_l[r]$ et $\overline{X_j}[r]$ sont respectivement les vecteurs de longueur w correspondant à la $r^{\text{ème}}$ ligne des matrices X_l et $\overline{X_j}$.

La matrice S_w, de taille $w \times w$, est donc la somme de Nh matrices de rang inférieur ou égal à un. Chacune de ces matrices est construite depuis au plus $(N-k)h$ observations linéairement indépendantes, car on a :

$$\forall\, j \in \{1, \dots, k\},\ \forall\, r \in \{1, \dots, h\}, \qquad \frac{1}{N_j}\sum_{X_l\in\Omega_j}X_l[r] = \overline{X_j[r]}$$

Le rang R de la matrice S_w vérifie donc :

$$R = \text{rang}(S_w) \leq \min(w, (N-k)h)$$

Par un raisonnement analogue, on peut montrer que :

$$R' = \text{rang}(\Sigma_w) \leq \min(h, (N-k)w)$$

Montrons maintenant que, dans le contexte de la reconnaissance automatique de visages, l'ADL2Do ne souffre pas du problème de la singularité. Supposons que les vecteurs-lignes (issus d'une même position r dans la matrice-image) des visages d'une même classe Ω_j soient indépendants et identiquement distribués selon une loi f_{jr} :

$$\forall\, X_l \in \Omega_j,\ \forall\, r = \{1, \ldots, h\}, \qquad X_l[r] \sim f_{jr}$$

Sous ces hypothèses, on obtient :

$$R = \text{rang}(Sw) = \min(w, (N-k)h) \tag{4.22}$$

et si l'on suppose vérifiées des conditions analogues sur les colonnes des matrices-images :

$$R' = \text{rang}(\Sigma_w) = \min(h, (N-k)w) \tag{4.23}$$

Or, puisque la construction du modèle nécessite que l'on ait au moins deux images par classe, on a forcément $N \geq 2k$ et par conséquent $N - k \geq k$. Dans le contexte de la reconnaissance de visages, on a généralement $w \ll kh$ et $h \ll kw$ et les conditions suivantes de non-singularité des matrices S_w et Σ_w sont donc vérifiées :

$$w < (N-k)h \quad \text{et} \quad h < (N-k)w \tag{4.24}$$

Par conséquent, l'ADL2Do peut être appliquée directement sur les visages, à la différence de l'ADL 1D. Afin d'illustrer cette notion, prenons l'exemple de la base ORL (*cf.* annexe A), couramment utilisée pour l'évaluation d'algorithmes de reconnaissance de visages. Celle-ci contient $k = 40$ personnes sous 10 vues différentes. Supposons que l'on retienne $N_j = 5$ vues par classe, pour chacune des 40 personnes représentées. On dispose donc de $N = 200$ images de visages, de taille $h \times w = 112 \times 92$. Comme nous l'avons vu au chapitre 3., il est impossible d'appliquer directement l'ADL sur cette base si celle-ci est modélisée de façon 1D, c.-à-d. par des vecteurs-images. En effet, dans ce cas, le rang de la matrice de covariance intra-classe associée (de taille $hw = 10304$) est inférieur à $\min(hw, N-k) = 160$. La matrice de covariance intra-classe est donc singulière. Aussi est-il nécessaire de mettre en œuvre l'une des variantes exposées en section 3.3, dont nous avons déjà évoqué les désavantages. Par contre, pour l'ADL2DoL, la matrice S_w, de taille $w \times w = 92 \times 92$, est de rang

$$R = \min(w, (N-k)h) = \min(92, 17920) = 92$$

Par conséquent, la matrice S_w est de rang plein et l'ADL2DoL est directement applicable. De la même manière, l'ADL2DoC peut s'utiliser directement, puisque la matrice Σ_w, de taille $h \times h = 112 \times 112$, est de rang $\min(h, (N-k)w) = \min(112, 14720) = 112$.

Nous venons donc de montrer qu'utiliser une modélisation 2Do pour appliquer l'ADL (algorithme d'ADL2Do) revient à travailler sur les lignes ou les colonnes des images et ainsi à augmenter artificiellement le nombre d'observations tout en diminuant leur dimensionnalité. Cela permet de contourner le problème de la singularité.

4.4.2.4 Interprétation géométrique et hypothèses nécessaires

Nous avons pour l'instant présenté des approches originales pour l'extraction de signature (à savoir l'ADL2DoL et l'ADL2DoC) sans réellement fournir une interprétation physique des signatures obtenues, ni étudier les conditions (portant sur les observations en entrée) sous lesquelles les classifieurs ainsi construits sont optimaux. Cette section vise à éclaircir ces deux points.

Réécrivons la matrice de covariance intra-classe généralisée S_w sous sa forme (4.21) :

$$S_w = \frac{1}{N}\sum_{j=1}^{k}\sum_{X_l\in\Omega_j}\sum_{r=1}^{h}(X_l[r]-\overline{X_j}[r])(X_l[r]-\overline{X_j}[r])^T$$

Les notations utilisées sont détaillées en section 4.4.2.3. On peut montrer que la matrice de covariance inter-classe généralisée S_b (donnée en équation (4.14)) peut s'écrire :

$$S_b = \frac{1}{N}\sum_{j=1}^{k}N_j\sum_{r=1}^{h}(\overline{X_j}[r]-\overline{X}[r])(\overline{X_j}[r]-\overline{X}[r])^T \tag{4.25}$$

où $\forall r=1,\ldots,h$, $\overline{X}[r]=\frac{1}{N}\sum_{j=1}^{k}\left(N_j\overline{X_j}[r]\right)$. Si l'on suppose que $\forall r\in\{1,\ldots,h\}$, $\forall j\in\{1,\ldots,k\}$, les vecteurs $X_l[r]$ issus de la classe Ω_j sont des observations indépendantes et identiquement distribuées selon une loi multinormale, de moyenne :

$$\mu^{(j,r)}=\frac{1}{N_j}\sum_{X_l\in\Omega_j}X_l[r]$$

et de matrice de covariance :

$$S^{(j,r)}=\frac{1}{N_j}\sum_{X_l\in\Omega_j}(X_l[r]-\mu^{(j,r)})(X_l[r]-\mu^{(j,r)})^T=\frac{1}{N_j}\sum_{X_l\in\Omega_j}(X_l[r]-\overline{X_j}[r])(X_l[r]-\overline{X_j}[r])^T$$

et que de plus les matrices de dispersion $S^{(j,r)}$ sont égales :

$$\forall\{j_1,j_2\}\in\{1,\ldots,k\}^2,\ \forall\{r_1,r_2\}\in\{1,\ldots,h\}^2,\ S^{(j_1,r_1)}=S^{(j_2,r_2)},$$

alors on peut considérer que l'ADL2DoL consiste à appliquer une ADL à kh classes, chaque classe $\Omega^{(j,r)}$ étant constituée des $r^{\text{ème}}$ lignes des images de Ω_j. Le classifieur ainsi obtenu porte donc sur la classification des lignes des images et l'on peut montrer qu'il est optimal au sens de la règle de Bayes (*cf.* section E.2 en p. 192 de l'annexe E). La table 4.4 donne une comparaison des caractéristiques de l'ADL2DoL et d'une procédure d'ADL standard qui serait appliquée directement aux vecteurs-images de la base d'apprentissage.

Nous avons vu en section 4.4 que l'ADL2Do permet d'éviter implicitement le problème de la singularité. Le critère de séparation est basé non pas sur k classes (identité des visages), mais sur kh classes, chacune correspondant à une identité et à une position verticale r fixée dans l'image. On cherche donc à séparer les lignes de l'image (pour une position r dans l'image fixée) correspondant à des identités Ω_j différentes. Les groupes Ω_j peuvent être vus comme des *méta-classes*, contenant l'ensemble des lignes associées. Par un raisonnement analogue, on trouve que l'ADL2DoC revient à effectuer une ADL à kw classes sur les colonnes des images de la matrice d'apprentissage et repose sur les mêmes méta-classes.

	ALD2DoL	**ADL Standard**
Nombre d'observations	Nh	N
Nombre de classes	kh	k
Nombre d'observations par classe	N_j	N_j
Taille des observations	w	wh
Singularité de $\mathbf{S}_w$ si	$Nh < w + kh$	$N < wh + k$

TAB. 4.4 – *Caractéristiques comparées de l'ADL2DoL et de l'ADL standard.*

Intéressons-nous maintenant à l'interprétation géométrique des signatures obtenues. Notons $P = [P_1, \ldots, P_g]$ la matrice de projection de l'ADL2DoL. La signature assignée à une image X par l'ADL2DoL est :

$$X^P = XP = \begin{bmatrix} X[1]^T P_1 & \ldots & X[1]^T P_i & \ldots & X[1]^T P_g \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ X[r]^T P_1 & \ldots & X[r]^T P_i & \ldots & X[r]^T P_g \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ X[h]^T P_1 & \ldots & X[h]^T P_i & \ldots & X[h]^T P_g \end{bmatrix}$$

Chaque élément $X^P[r,i]$ de la signature X^P est donc la projection de la $r^{\text{ème}}$ ligne de l'image X sur le $i^{\text{ème}}$ vecteur discriminant de l'ADL2DoL. Par un raisonnement similaire, on trouve que chaque élément $X^Q[i,c]$ de la signature X^Q (obtenue par ADL2DoC) est la projection de la $r^{\text{ème}}$ colonne de l'image X sur le $i^{\text{ème}}$ vecteur discriminant de l'ADL2DoC.

4.4.3 Classification

Dans cette section, nous cherchons à définir le mode de classification (mesure de dissimilarité et règle d'affectation) le plus efficace dans le contexte de l'ADL2Do.

4.4.3.1 Approche proposée

Lorsqu'un visage-requête X doit être reconnu, on commence par calculer sa signature X' associée ($X' = X^P$ pour l'ADL2DoL ou $X' = X^Q$ pour l'ADL2DoC). Puis, on calcule la distance entre cette signature et celles (X'_l) de la base de connaissance, selon l'une des mesures de dissimilarité détaillées ci-après, de manière à assigner une identité Ω_j^* à ce visage-requête. La règle de décision est soit au plus proche voisin (déjà évoquée en p. 89) :

$$\Omega_j^* = \underset{j=1,\ldots,k}{\text{Argmin}} \left[\min_{X_l \in \Omega_j} d(X', X'_l) \right] \tag{4.26}$$

soit à la plus proche moyenne :

$$\Omega_j^* = \underset{j=1,\ldots,k}{\text{Argmin}} \left[d(X', \overline{X'_j}) \right] \tag{4.27}$$

où d est une mesure de dissimilarité entre matrices.

Dans cette section, nous testons l'efficacité de différentes mesures de dissimilarité pour la classification des signatures issues de l'ADL2Do. Prenons l'exemple de l'ADL2DoL. Les signatures $X^P = [X_1^P, \ldots, X_g^P]$ sont de taille $h \times g$. Notons $D_{L_p}^{(P)}$ les distances suivantes entre deux matrices-signatures X^P et Y^P, qui généralisent les distances de Minkowski entre vecteurs (*cf.* annexe D) :

$$D_{L_p}^{(P)}(X^P, Y^P) = \sum_{i=1}^{g} d_{L_p}(X_i^P, Y_i^P) \tag{4.28}$$

où $D_{L_p}^{(P)}$ est la somme des distances de Minkowski d_{L_p} entre vecteurs-colonnes des signatures. Pour la comparaison de signatures issues de l'ADL2DoC, on appliquera cette distance sur les signatures transposées de manière à définir la distance $D_{L_p}^{(Q)}$ comme somme des distances de Minkowski entre vecteurs-lignes des signatures :

$$D_{L_p}^{(Q)}(X^Q, Y^Q) = \sum_{i=1}^{g} d_{L_p}(X_i^{Q^T}, Y_i^{Q^T}) \tag{4.29}$$

Dans la suite de cette thèse, nous nous référerons à ces distances $D_{L_p}^{(P)}$ et $D_{L_p}^{(Q)}$ par le sigle D_{L_p}.

Pour $p = 1$, D_{L_p} est nommée *distance de Manhattan* ; pour $p = 2$ il s'agit de la *distance Euclidienne*. On étend cette définition à des mesures de dissimilarité fractionnaires, avec $p \in]0, 1[$. Il ne s'agit pas de distances, car l'inégalité triangulaire n'est pas vérifiée. Leurs boules unités associées sont illustrées en figure D.1 de l'annexe D. Ces mesures fractionnaires sont réputées efficaces pour la classification de données de grandes dimensions [AHK01, LDGF04]. On peut donc attendre des bonnes performances sur notre problème, puisque les dimensions des matrices-signatures sont assez importantes.

4.4.3.2 Choix de la mesure de dissimilarité

Afin de déterminer la mesure de dissimilarité la plus adaptée à la reconnaissance de visages, nous avons mené des expérimentations sur la base ORL (*cf.* annexe A). La base d'apprentissage (servant également de base de connaissance) est constituée de 5 images sélectionnées aléatoirement, pour chacune des 40 personnes enregistrées (soit un total de $N = 200$ images d'apprentissage), à une résolution suffisante de 61×46 pixels. L'impact de la résolution sur les performances du système sera étudié en section 4.4.5.1. La base de test est constituée des 5 images par personne restantes (soit un total de 200 visages-requêtes). Cette opération est répétée cinq fois. Pour chacune des deux versions de l'ADL2Do et chaque mesure de dissimilarité (avec une règle au plus proche voisin ou à la plus proche moyenne), on calcule le taux de reconnaissance moyen (sur les cinq partitions) en fonction du nombre g de vecteurs propres retenus.

Les figures 4.5-(a-d) fournissent une comparaison des taux de reconnaissance moyens de l'ADL2DoL et de l'ADL2DoC, utilisées conjointement avec mesures de dissimilarité D_{L_p}, pour p allant de 0,3 à 2. Les figures 4.5-(a-b) sont construites en utilisant une stratégie d'affectation au plus proche voisin, tandis que les taux de reconnaissance des figures 4.5-(c-d) sont estimés à l'aide de la plus proche moyenne. L'étude de ces graphes nous apporte plusieurs enseignements.

Tout d'abord, la règle d'affectation à la plus proche moyenne est moins efficace que le plus proche voisin, ceci quelle que soit la mesure de dissimilarité considérée. Cela provient du fait que les données sont trop dispersées dans l'espace de projection pour que l'étude des moyennes des

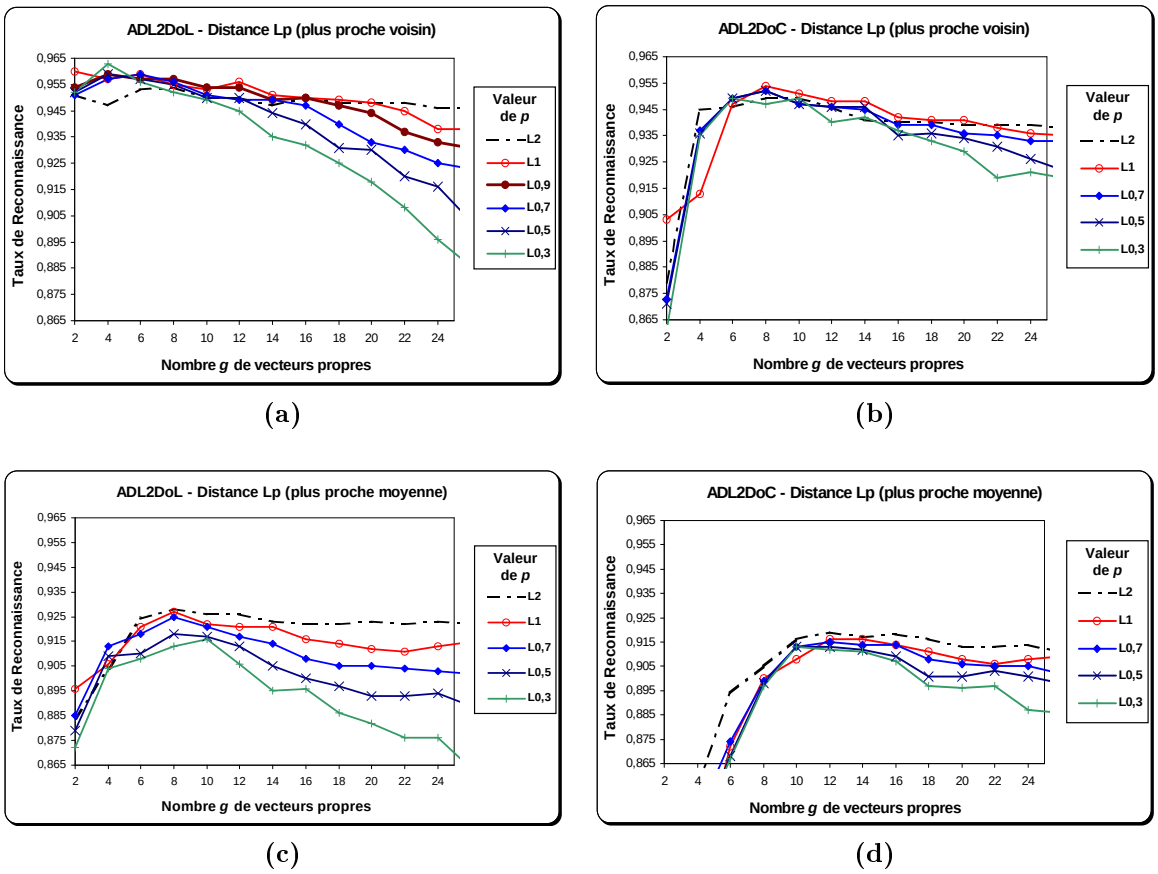


FIG. 4.5 – *Impact du paramètre p de la mesure de dissimilarité D_{L_p} sur les taux de reconnaissance moyens calculés sur cinq partitions de la base ORL (illustrée en figure A.3, p. 172). Les techniques évaluées en (a) et (b) sont respectivement l'ADL2DoL et l'ADL2DoC avec une stratégie au plus proche voisin. Les figures (c) et (d) sont construites en utilisant respectivement l'ADL2DoL et l'ADL2DoC, l'affectation étant réalisée selon la plus proche moyenne.*

classes suffise à garantir d'excellents taux de reconnaissance. On retrouve ce phénomène avec la plupart des techniques statistiques de projection (*cf.* section 2.2.2).

Nous remarquons également que, quelle que soit la mesure de dissimilarité utilisée, le fait de retenir un nombre trop important de vecteurs propres dans le modèle engendre une baisse des taux de reconnaissance. Nous reviendrons sur ce point en section 4.4.4. Pour l'ADL2DoL, avec une règle d'affectation au plus proche voisin et un faible nombre g de vecteurs propres, plus le paramètre p est petit, plus les performances sont élevées (*cf.* figure 4.5-a). Cependant, les écarts entre les maxima respectifs, de l'ordre de 0,5% sont peu significatifs. De plus, il semblerait que, plus la valeur de p est petite, plus les performances sont dépendantes du nombre de vecteurs propres retenus dans le système : plus p est faible, plus les taux de reconnaissance chutent rapidement lorsque la valeur de g augmente. Pour l'ADL2DoC, on observe le même phénomène avec des écarts encore moins significatifs (voir figure 4.5-b). En revanche, avec une stratégie d'affectation à la plus proche moyenne, la distance D_{L_2} donne de meilleurs résultats que les mesures de dissimilarité D_{L_p} où $p < 2$.

Nous venons d'observer que les mesures fractionnaires peuvent donner des résultats légère-

ment meilleurs que les distances de Minkowski usuelles, mais que leurs performances sont très influencées par le nombre de vecteurs propres retenus dans le modèle. D'autres expérimentations (non fournies ici pour des raisons de place) montrent que la distance D_{L_2} est la plus invariante à tout un ensemble de facteurs influents, tels que la taille de la base. Puisque nous avons de plus montré dans cette section que la règle d'affectation au plus proche voisin est plus efficace que la règle à la plus proche moyenne, nous utiliserons pour nos expérimentations la distance Euclidienne D_{L_2} au plus proche voisin.

4.4.4 Sélection du nombre de composantes

Nous venons de voir que le nombre g de vecteurs propres retenus dans le modèle a une forte influence sur les performances de l'ADL2Do. En effet, le fait de retenir un nombre trop important de vecteurs propres engendre une baisse des taux de reconnaissance. Cela peut être expliqué par un phénomène de *surapprentissage*. En effet, le paramètre g détermine la quantité d'information provenant de la base d'apprentissage et utilisée pour la classification des données de test. Le fait de retenir un volume trop important d'information peut conduire à la prise en compte du bruit de la base d'apprentissage, ce qui nuit à la capacité de généralisation du système. Si l'on ajoute à cela le fait que, plus on retient de vecteurs propres, plus la phase de classification est coûteuse en termes de nombre de calculs, on comprend qu'il est nécessaire de définir des stratégies permettant de sélectionner le nombre optimal de vecteurs propres à retenir, c.-à-d. la valeur g^* du paramètre g qui fournit le meilleur compromis entre performance du système et coût de classification.

4.4.4.1 Le critère du Lambda de Wilks

Un premier mode de sélection repose sur l'utilisation du *critère du Lambda de Wilks* [Sap90]. Il s'agit de l'un des tests couramment mis en œuvre pour la sélection de composantes dans le cadre de modèles d'analyse de variance, au même titre que le test de la *trace de Pillai*, de la *trace de Lawley-Hotelling* et de *la plus grande racine de Roy* [Sap90]. Nous proposons de le mettre en œuvre dans le cadre d'une méthodologie de sélection séquentielle des vecteurs propres à retenir. Considérons le cas de l'ADL2DoL. Notons S_T la matrice de dispersion totale généralisée, obtenue en sommant les dispersions intra- et inter-classe :

$$S_T = S_w + S_b \tag{4.30}$$

Le critère du Lambda de Wilks permet de tester le pouvoir discriminant des derniers vecteurs propres de la matrice $S_w^{-1}S_b$ afin de déterminer s'il est judicieux ou non de les rejeter du modèle. Il repose sur la définition des matrices de Wishart et de la loi de Wilks.

Définition 4.1 *Une matrice M, de taille $p \times p$, a une distribution de Wishart $W_p(n, \Sigma)$ si M peut s'écrire $M = X^T X$, où X est une matrice aléatoire de taille $n \times p$, définie de la façon suivante : les n lignes de X sont des vecteurs aléatoires de même loi $\mathcal{N}(0, \Sigma)$ indépendantes.*

Définition 4.2 *Soient A et B deux matrices de Wishart $W_p(m, \Sigma)$ et $W_p(n, \Sigma)$ indépendantes, où $m \geq n$, alors le quotient :*

$$\Lambda = \frac{|A|}{|A+B|} \tag{4.31}$$

a une distribution de Wilks de paramètres p,m,n, notée $\Lambda(p,m,n)$.

est grand, alors on peut approximer la loi de Wilks par l'approximation de Bartlett :

$$-\left[m-\frac{1}{2}(p-n+1)\right]\ln\Lambda(p,m,n) \approx \chi^2_{np} \tag{4.32}$$

Posons comme hypothèse nulle du test H_0 : les $w-g$ derniers vecteurs propres de la matrice $S_w^{-1}S_b$ n'ont aucun pouvoir discriminant. L'hypothèse alternative H_1 est donc que le sous-espace engendré par ces $w-g$ derniers vecteurs propres a un pouvoir discriminant. Supposons l'hypothèse H_0 vérifiée. Regroupons les $w-g$ derniers vecteurs dans la matrice notée $P_\perp = [P_{g+1}, P_{g+2}, \ldots, P_w]$. Puisque nous faisons l'hypothèse que les observations sont Gaussiennes et que toutes les classes ont même variance (voir section 4.4.2.4), dire que les vecteurs contenus dans $P_\perp$ sont non discriminants (H_0) revient à dire que les centroïdes des classes projetées sur $P_\perp$ sont confondus :

$$\overline{X_1}P_\perp = \overline{X_2}P_\perp = \ldots = \overline{X_k}P_\perp \tag{4.33}$$

Sous cette hypothèse, il est facile de montrer que les matrices de covariance généralisées $S_w^{P_\perp}$ et $S_b^{P_\perp}$ des données projetées sur $P_\perp$ sont des matrices de Wishart de taille $(w-g)\times(w-g)$, respectivement de degrés de liberté $(N-k)h$ et $(k-1)h$. Par conséquent, si l'on note λ_i la valeur propre associée au vecteur propre P_i, le quotient :

$$\Lambda' = \frac{|S_w^{P_\perp}|}{|S_T^{P_\perp}|} = \frac{|S_w^{P_\perp}|}{|S_w^{P_\perp}+S_b^{P_\perp}|} = \frac{1}{|{P_\perp}^T S_w^{-1} S_b P_\perp + I_{w-g}|} = \prod_{i=g+1}^{w}\frac{1}{1+\lambda_i} \tag{4.34}$$

suit la loi de Wilks $\Lambda(w-g,(N-k)h,(k-1)h)$ et l'on a (approximation de Bartlett) :

$$-\left[Nh-\frac{1}{2}(w-g+(k+1)h+1)\right]\ln\prod_{i=g+1}^{w}\frac{1}{1+\lambda_i} \approx \chi^2_{(w-g)(k-1)h} \tag{4.35}$$

Par un raisonnement analogue, on obtient pour l'ADL2DoC :

$$-\left[Nw-\frac{1}{2}(h-g+(k+1)w+1)\right]\ln\prod_{i=g+1}^{h}\frac{1}{1+\lambda_i} \approx \chi^2_{(h-g)(k-1)w} \tag{4.36}$$

Ce test peut être utilisé de manière ascendante ou descendante, selon que l'on choisit d'ajouter ou de retirer de manière séquentielle des vecteurs propres du modèle. Les résultats expérimentaux ayant prouvé qu'un faible nombre g de composantes est généralement suffisant pour obtenir d'excellents taux de reconnaissance, la méthode ascendante est moins coûteuse et nous la préférerons en général. Dans le contexte d'une sélection ascendante, si la *p-valeur* [14] p est inférieure au seuil $\alpha = 0{,}05$, on rejette H_0, on augmente d'un le nombre g de vecteurs propres ($g := g+1$), puis on réitère ce test, et ainsi de suite jusqu'à déterminer la valeur de g^* pour laquelle $p > \alpha$. On considère que chaque test est effectué de manière indépendante, puisque nous n'incorporons pas dans l'hypothèse à tester le fait que ce test a déjà été effectué avec une autre valeur de g. Pour une sélection descendante au contraire, tant que $p > \alpha$, on ne peut rejeter H_0 au niveau de signification α : on diminue d'un le nombre g de vecteurs propres ($g := g-1$) et on réitère ce test jusqu'à déterminer la valeur de g^* telle que, pour g^*-1, la p-valeur p est inférieure à α.

14. La p-valeur, notée p, est la probabilité sous l'hypothèse H_0 que la statistique de test prenne une valeur au moins aussi extrême que celle observée dans l'expérience. L'*erreur de première espèce* consiste à rejeter H_0 à tort. Afin de contrôler le risque d'erreur de première espèce, on fixe une probabilité d'erreur maximale, appelée *seuil* et notée α. On rejette H_0 au niveau de confiance α si et seulement si la p-valeur p est inférieure au seuil α.

Dans la pratique, le nombre d de degrés de liberté (respectivement $d = (w-g)(k-1)h$ pour l'ADL2DoL et $d = (h-g)(k-1)w$ pour l'ADL2DoC) de la loi du χ^2, qui dépend de la résolution des images, peut être très grand. Généralement, les tables des fractiles de la loi du χ^2 ne sont disponibles que pour des valeurs de d inférieures ou égales à 100. Même s'il existe des approximations pour $d > 100$, on préférera éviter de mener des tests avec un nombre de degrés de liberté trop importants, d'autant plus que généralement un faible nombre de composantes g (inférieur à 20) suffit. On ne testera donc pas en général le pouvoir de séparation de l'ensemble des vecteurs propres P_i, pour i allant de $g+1$ à w. On se restreindra à l'étude des $g' - g$ derniers vecteurs propres, où $g' < w$. Toujours dans cette optique de restreindre la valeur de d, on préférera utiliser ce test pour des images de résolution la plus réduite possible.

Ce mode de sélection est rapide et son efficacité est prouvée (il est notamment utilisé dans les logiciels de statistiques les plus prisés tels que SAS, Splus et STATISTICA). Néanmoins, le critère du Lambda de Wilks repose sur la séparation des classes de la base d'apprentissage et non sur des résultats de classification estimés sur des bases de test indépendantes. Or, notre but est de concevoir un système doté de la meilleure capacité de généralisation possible. Nous proposons donc d'autres modes de sélection, basés sur l'estimation de la capacité de généralisation.

4.4.4.2 Utilisation d'une base de validation

La première approche proposée consiste à utiliser une base de validation pour déterminer g^*. Plaçons-nous dans le contexte de l'ADL2DoL. Les résultats qui découlent de l'analyse qui suit pourront directement être étendus au cas de l'ADL2DoC. On calcule, à partir de la base d'apprentissage, la matrice P de projection de l'ADL2DoL constituée de tous les w vecteurs propres de la matrice $S_w^{-1}S_b$. Puis, on considère les w sous-matrices $P^{(g)}$ de P, où $P^{(g)} = [P_1, P_2, \ldots P_g]$ (où les P_i sont les vecteurs propres rangés dans l'ordre de leurs valeurs propres décroissantes), pour g allant de 1 à w, chacune de ces sous-matrices permettant de définir un classifieur différent. La sélection du meilleur classifieur se fait à l'aide d'une base de validation distincte de la base d'apprentissage et contenant N' images de visages enregistrés dans cette dernière. On calcule le taux de reconnaissance obtenu par chacun de ces classifieurs sur la base de validation (contenant N' images) et on retient le g^* optimal, c.-à-d. celui associé au classifieur $P^{(g^*)}$ fournissant le meilleur compromis entre taille des signatures et taux de reconnaissance. Dans la pratique, il est inutile de tester l'ensemble des g classifieurs possibles : les résultats expérimentaux montrent que l'on peut restreindre la recherche à $g \leq 20$. Néanmoins, ce mode de sélection est coûteux, puisqu'il nécessite, pour l'évaluation de chaque classifieur, N' assignations au plus proche voisin parmi une base d'apprentissage contenant N images. De plus, il n'est pas toujours possible de disposer de suffisamment de données pour constituer une base de validation distincte de la base d'apprentissage.

Une solution basée sur l'estimation de la capacité de généralisation mais ne nécessitant pas de disposer d'exemples supplémentaires en plus de ceux de la base d'apprentissage consiste à créer artificiellement un ensemble de validation, par modification numérique des images de la base d'apprentissage. On pourra pour cela utiliser des images « miroir » (voir section 3.3.3.1), et/ou appliquer des artefacts comme nous l'avons fait pour tester la robustesse de l'ACP2D en section 4.3.4. Néanmoins, si les modifications sont insuffisantes, cette technique peut engendrer un surapprentissage des données connues.

4.4.4.3 Discussion

Nous avons présenté dans les sections 4.4.4.1 et 4.4.4.2 deux modes de sélection du nombre de vecteurs propres du modèle, selon que l'estimation du pouvoir discriminant est effectuée à partir de la base d'apprentissage ou d'une base de visages indépendante de celle-ci. Afin de maximiser la capacité de généralisation du modèle, nous préférerons généralement utiliser une base de validation indépendante des bases d'apprentissage et de test, si nous disposons de suffisamment de données pour ce faire. Dans le cas contraire, nous mettrons en œuvre le critère du Lambda de Wilks.

4.4.5 Impact de différents facteurs sur les performances du système

Outre la mesure de dissimilarité utilisée et le nombre de vecteurs propres retenus, trois autres facteurs peuvent avoir un impact important sur les performances de l'ADL2Do. Puisqu'il s'agit d'une méthode globale, la résolution des images peut influer de manière significative sur les performances. Le nombre de classes (nombre de personnes enregistrées dans la base) et le nombre d'images par classe sont également deux facteurs potentiellement très influents.

4.4.5.1 Impact de la résolution des images

Cette section vise à évaluer l'influence de la résolution des images sur les performances du système, afin d'en déduire une résolution optimale. Pour cela, nous utilisons un sous-ensemble de la base PF01 contenant la totalité des 107 personnes enregistrées et contenant des variations de pose en profondeur (de type hochement de la tête ou négation). Les images sont normalisées conformément au processus détaillé en annexe C pour obtenir des images de visages de taille 150×130 pixels (les positions des yeux sont fournies avec la base). Les bases utilisées pour l'apprentissage et le test sont illustrées en figure 4.9-a, p. 113. La base d'apprentissage contient 535 images de visages, c.-à-d. cinq vues par personne, sous des poses différentes mais presque frontales. La base de test contient 428 images, c.-à-d. quatre vues par personne, sous des poses plus éloignées de la pose frontale. Les conditions d'illumination ainsi que les expressions faciales sont similaires dans les bases d'apprentissage et de test. Pour chaque résolution, la base de test est comparée à la base d'apprentissage selon la distance Euclidienne au plus proche voisin. La figure 4.6 montre les taux de reconnaissance ainsi obtenus, à différentes résolutions. On s'aperçoit qu'une diminution de la résolution des images, jusqu'à un certain point, n'engendre pas de baisse significative des taux de reconnaissance. On remarque qu'une résolution comprise entre 45×39 et 75×65 pixels fournit un bon compromis entre coût calculatoire de classification et performances. Il en est de même pour d'autres expérimentations, menées sur différentes bases de visages. Par conséquent, nous choisissons de travailler avec des images de résolution comprise entre ces deux valeurs.

4.4.5.2 Impact du nombre de personnes enregistrées

Afin d'étudier l'impact du nombre k de classes d'apprentissage sur les performances du modèle, nous avons mené des expérimentations sur la base ORL et sur la sous-base de PF01 utilisée en section 4.4.5.1. La base ORL sert à évaluer l'impact du nombre de classes lorsque celui-ci est inférieur à 40, tandis que la base PF01 est utilisée pour des valeurs de k allant de 50 à 107.

Détaillons le protocole expérimental utilisé. La base ORL (*cf.* figure A.3, p. 172) contient initialement 10 vues de 40 personnes, redimensionnées à une taille suffisante de 61×46 pixels. On

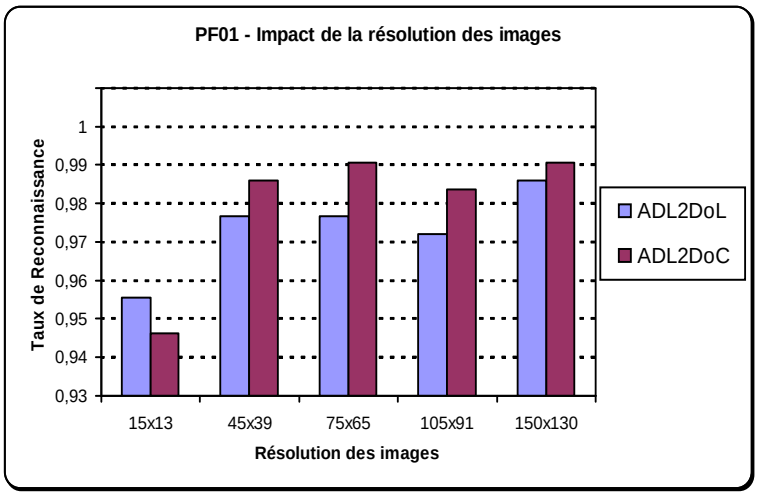
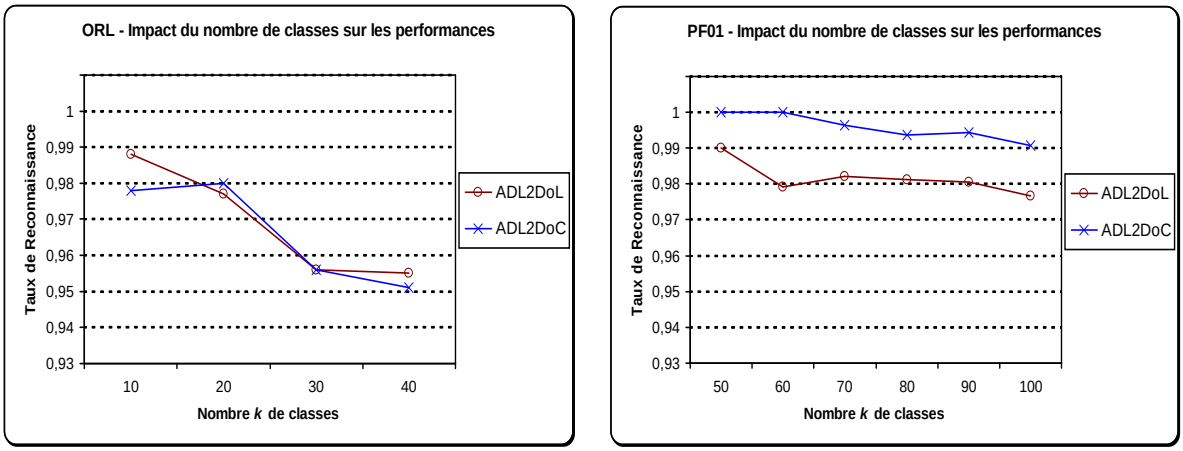


FIG. 4.6 – *Taux de reconnaissance comparés de l'ADL2DoL et de l'ADL2DoC sur un sous-ensemble de la base PF01, à différentes résolutions.*

effectue un double tirage aléatoire : dans un premier temps, on sélectionne au hasard k identités parmi les 40 personnes représentées. Dans un second temps on tire au hasard, pour chacune de ces k personnes, 5 vues parmi les 10 vues disponibles. Celles-ci serviront à l'apprentissage. Les vues restantes des k personnes enregistrées constituent la base de test. Ce double tirage aléatoire est répété 10 fois. Les taux de reconnaissance moyens sur les 10 partitions, pour des valeurs de k comprises entre 10 et 40, sont donnés dans la figure 4.7-a. Une stratégie similaire est appliquée sur la sous-base de PF01 utilisée précédemment (les images étant normalisées et redimensionnées à une taille de 75×65 pixels). On obtient ainsi les taux de reconnaissance montrés en figure 4.7-b, pour un nombre de classes variant de 50 à 107.



(a) **(b)**

FIG. 4.7 – *Impact de la taille sur les performances observées sur les taux de reconnaissance de l'ADL2DoL et de l'ADL2DoC évalués (a) sur la base ORL et (b) sur la base PF01.*

Dans les deux cas, on constate une baisse des taux de reconnaissance lorsque le nombre de classes augmente. Notons que le nombre d'images par classe est le même pour les deux bases. Cependant, la valeur de la pente descendante est différente : les taux de reconnaissance calculés

sur la base PF01 chutent moins vite que sur la base ORL. Les premiers étant globalement supérieurs aux seconds, il semblerait que la valeur de la pente descendante soit proportionnelle à la performance du système. Celle-ci est directement liée au nombre et à l'amplitude des variations (la base ORL contient plus de sources de variation que la sous-base de PF01 considérée). Ainsi, plus les images sont bruitées, plus l'ajout d'une personne supplémentaire accroît la complexité de l'espace des visages et plus il est difficile de reconnaître ces visages. La figure 4.7 montre que, dans des conditions de prise de vue contrôlées et malgré des changements de pose relativement importants, l'ADL2Do permet de reconnaître à 99,1% les visages d'une centaine de personnes.

4.4.5.3 Impact du nombre d'exemples par classe

Cette section vise à étudier l'impact du nombre d'exemples connus par classe sur les performances du système. Pour cela, nous utilisons la sous-base de PF01 décrite en section 4.4.5.1. Le nombre k de classes est fixé à 107. Le nombre de vues par classe est le même pour toutes les classes, c.-à-d. $\forall j = 1, \ldots k$, $N_j = n$. On fait varier n de 2 à 8. Pour $n < 5$, la base de test est constituée des mêmes images que pour $n = 5$ (*cf.* section 4.4.5.1) et on effectue un tirage aléatoire pour déterminer les vues à inclure dans la base d'apprentissage. Pour $N_j > 5$, on procède à un tirage aléatoire sur la base de test pour choisir les vues à transférer de la base de test vers la base d'apprentissage. De cette manière, aucun recoupement n'est possible entre base d'apprentissage et de test. Afin de favoriser l'homoscédasticité, on regroupe les vues des 107 personnes prises dans des mêmes conditions dans la même base (apprentissage ou test). Les taux de reconnaissance en fonction du nombre n d'exemples par classe sont donnés en figure 4.8.

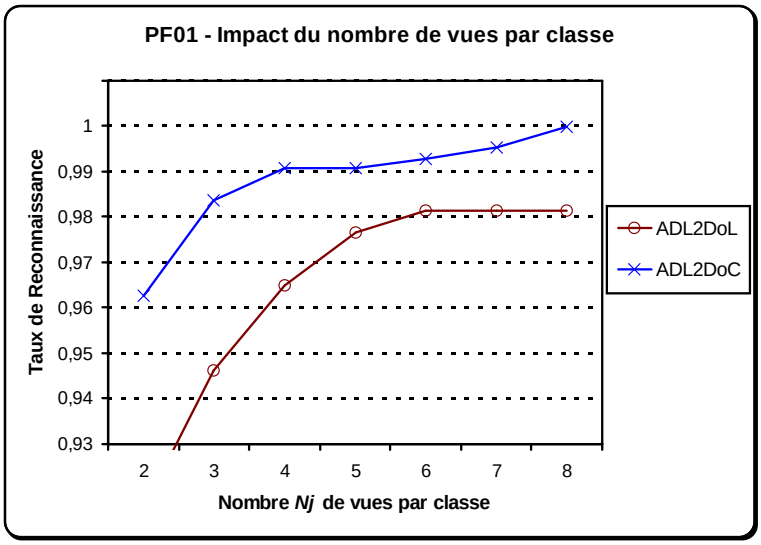


FIG. 4.8 – *Taux de reconnaissance comparés de l'ADB, de l'ADL2DoL et de l'ADL2DoC sur un sous-ensemble de la base PF01, lorsque le nombre N_j de vues par classe varie.*

Cette figure met en lumière le fait que le facteur n influe sur les performances du système. On peut remarquer qu'il est dans tous les cas préférable de disposer d'un nombre d'exemples par classe important. Tandis qu'une valeur de $n = 5$ suffit à garantir des performances optimales pour l'ADL2DoC, le point d'inflexion de la courbe croissante de l'ADL2DoL se situe plutôt en $n = 6$.

4.4.6 Évaluation des performances et comparaison aux techniques usuelles de projection statistique

Comparons maintenant les performances de l'ADL2Do avec celles d'autres techniques basées sur la projection statistique dans un contexte d'identification en monde fermé. Nous utilisons pour cela quatre expérimentations. La première, menée sur la base ORL, vise à comparer les performances de l'ADL2Do à celles de l'ACP2D et des principales techniques d'ADL 1D en présence de multiples sources de variation d'amplitude réduite (variations dans les conditions d'illumination, les poses de la tête et les expressions faciales). Les deuxième et troisième expérimentations, menées sur la base PF01, visent à étudier indépendamment l'influence de variations de pose et d'expression sur les performances de l'ADL2Do, de l'ACP2D et des *fisherfaces*. Notons que la troisième expérimentation vise également à étudier l'impact de la non-homoscédasticité des données en entrée dans un contexte d'aide à la décision (*cf.* section 1.7). La quatrième expérimentation, utilisant les bases FERET et BioId (*cf.* annexe A), fournit une comparaison des capacités de généralisation (à des personnes non enregistrées dans la base d'apprentissage) de ces différentes méthodes. Tandis qu'un extrait de la base ORL est donné en figure A.3 (p. 172), les bases d'apprentissage et de test utilisées pour les trois dernières expérimentations sont illustrées en figure 4.9 ci-après. Il est à noter qu'une étude des performances de l'ADL2Do en présence de variations dans les conditions d'illumination sera menée en section 5.2.5.3.

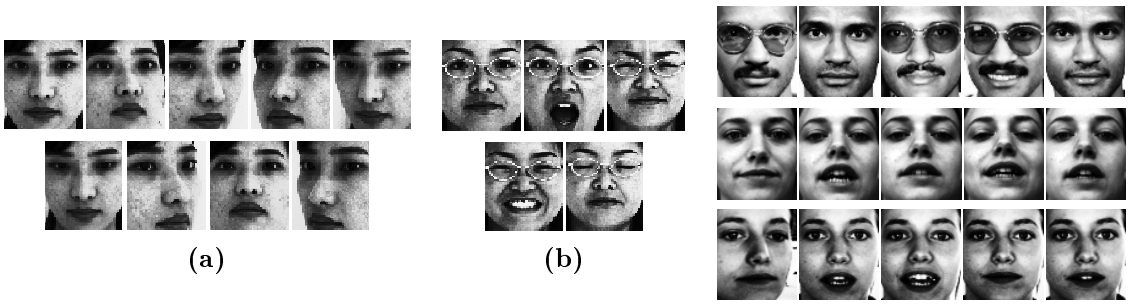


FIG. 4.9 – *Extraits des bases utilisées pour (a) la deuxième et (b) la troisième expérimentation : dans chaque cas, les images de la première ligne proviennent de la base d'apprentissage (également utilisée comme base de connaissance) et les images de la seconde ligne sont extraites de la base de test. Les images en première ligne du (c) proviennent de la base d'apprentissage de la quatrième expérimentation (images extraites de FERET). Les deuxième et troisième lignes correspondent respectivement à des extraits de la base de connaissance et de la base de test de cette dernière expérience (images extraites de BioId). Remarquons que les variations entre la base de connaissance et la base de test sont importantes (c'est la même personne qui est montrée en deuxième et troisième lignes de la colonne (c)).*

4.4.6.1 Expérimentation menée sur la base ORL

La table 4.5 fournit sur la base ORL une comparaison des taux de reconnaissance de l'ADL2Do et des principales techniques de projections statistique utilisées dans le contexte de la reconnaissance de visages. Le protocole expérimental est le suivant. La base contient 40 visages avec dix images par classe. Cette base est aléatoirement divisée en une base d'apprentissage et une base de test, chacune contenant cinq images par personne. Un taux de reconnaissance est calculé par

comparaison de la base de test à la base d'apprentissage. Cette opération est répétée p fois. Le taux de reconnaissance moyen sur ces p partitions est donné dans la table 4.5. Les résultats de l'ADL2Do et de l'ACP2D ont été obtenus par nos propres expérimentations, tandis que pour les autres techniques elles sont tirées des articles « Source », qui utilisent le même protocole expérimental détaillé ci-dessus. Notons que le nombre de vecteurs propres fournissant les meilleurs taux de reconnaissance sont respectivement $g = 5$, $g = 10$ et $g = 8$ pour l'ADL2DoL, l'ADL2DoC et l'ACP2D. Ce tableau met en lumière le fait que les performances de l'ADL2Do

Méthode	**Source**	**p**	**Taux de reconnaissance**
ADL2DoL	–	10	95,55%
ADL2DoC	–	10	95,05%
ACP2D	–	10	95%
ACP+ADL$_0$	[HLLM02]	50	95,63%
ADLRD	[DY03]	4	95,25%
fisherfaces	[HLLM02]	50	94,19%
ADL Directe	[YY01]	10	90,8%

TAB. 4.5 – *Taux de reconnaissance comparés de diverses techniques sur p partitions aléatoires de la base ORL. Les méthodes testées sont: l'ADL2DoL, l'ADL2DoC, l'ACP2D [YZFY04], l'algorithme d'ADL dans le noyau de Huang et al. [HLLM02] (ACP+ADL$_0$), l'ADL Régularisée de Dai et al. [DY03], l'ADL Directe (ADLD) [YY01] ainsi que la technique des fisherfaces [BHK97].*

(et surtout de l'ADL2DoL) sont comparables aux meilleures techniques issues de l'état de l'art. En effet, la technique reportant les meilleurs taux de reconnaissance (ACP+ADL$_0$) n'obtient qu'une amélioration non significative de 0,08% par rapport à l'ADL2DoL.

4.4.6.2 Expérimentations menées sur la base PF01

Pour la première expérimentation, nous utilisons les mêmes bases d'apprentissage et de test que celles détaillées en section 4.4.5.1 et présentent des variations en profondeur de la pose de la tête pouvant aller jusqu'à 45° (*cf.* figure 4.9-a). Les taux de reconnaissance comparés de l'ADL2DoL, de l'ADL2DoC et de l'ACP2D sont fournis dans la table 4.6. Cette table montre que l'ADL2Do (surtout sa version orientée en colonnes ADL2DoC), donne des résultats sensiblement meilleurs que les autres techniques issues de l'état de l'art (avec p. ex. 424 bonnes classifications sur 428 requêtes pour l'ADL2DoC contre 412 pour l'ACP2D). Il semblerait donc qu'elle soit plus tolérante que celles-ci à des variations de pose de la tête en profondeur.

La deuxième expérimentation menée sur la base PF01 utilise les bases d'apprentissage illustrées en figure 4.9-b. Toutes les vues utilisées présentent des conditions d'illumination similaires et une pose frontale. Seule l'expression faciale varie. La base d'apprentissage contient trois vues

Méthode	ADL2DoC	ADL2DoL	ACP2D	*fisherfaces*
Taux de reconnaissance	99,1%	97,7%	97%	96,3%

TAB. 4.6 – *Taux de reconnaissance comparés de l'ADL2DoC, de l'ADL2DoL, de l'ACP2D [YZFY04] et de l'algorithme des fisherfaces [BHK97] sur la sous-base de PF01 contenant des variations de pose.*

par personne pour chacune des 107 personnes présentes, soit au total 321 vues. Dans le cadre de certaines applications, l'hypothèse d'homoscédasticité n'est pas vérifiée : en effet, il n'est pas toujours possible de se procurer des vues prises dans les mêmes conditions pour chaque personne de la base. Afin de tester la robustesse à l'hypothèse d'homoscédasticité, les vues ne sont pas consistantes d'une personne à l'autre : pour chaque classe, la base d'apprentissage contient la vue avec une expression faciale neutre, ainsi que deux vues sélectionnées aléatoirement parmi les quatre expressions faciales proposées dans la base (voir figure 4.9-b). Étant donné que cette hypothèse est *a priori* nécessaire pour l'ADL2Do et les *fisherfaces*, mais pas pour l'ACP2D, on pourrait s'attendre à ce que, dans ces conditions, l'ACP2D donne de meilleurs résultats que les deux autres méthodes. La base de test contient les deux vues non sélectionnées de chaque classe. La base de test est comparée à la base d'apprentissage. Les résultats de classification sont analysés au travers d'une courbe *Cumulative Match Characteristic* (CMC), donnée en figure 4.10. Ce type de courbes, décrit en section 1.7.1 du chapitre 1, vise à évaluer les performances d'un système d'aide à la décision (voir section 1.7). Un visage est reconnu au rang r si une vue du même visage est parmi ses r plus proches voisins, au sens de la distance Euclidienne. C'est l'ADL2DoC qui est la plus performante des deux versions de l'ADL2Do.

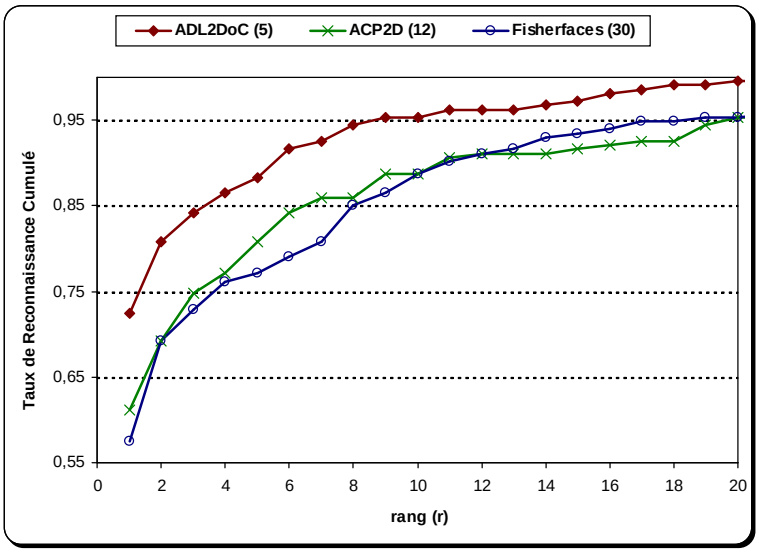


FIG. 4.10 – *Comparaison des courbes CMC de l'ADL2DoC, de l'ACP2D et des fisherfaces, sur la sous-base de PF01 contenant des variations de pose. Un visage est reconnu au rang r si une vue du même visage est parmi ses r plus proches voisins, au sens de la distance Euclidienne.*

Intéressons-nous tout d'abord aux taux de reconnaissance obtenus au rang $r = 1$. La figure 4.10 montre que l'ADL2DoC donne de bien meilleurs taux de reconnaissance que l'ACP2D

et les *fisherfaces*. Nous pouvons constater que les effets cumulés du faible nombre d'exemples par classe et de la non-homoscédasticité des données semblent plus affecter les *fisherfaces* que l'ADL2Do. Nous avons observé qu'environ 68% des mauvaises classifications de l'ACP2D correspondent à la mise en correspondance de personnes différentes mais avec la même expression faciale. En revanche, ce cas de figure ne représente que 52% des erreurs de l'ADL2Do, qui sont elles-mêmes moins nombreuses. En conséquence, on peut en déduire que l'ADL2DoC est plus robuste aux changements dans les expressions faciales et ce, malgré la non-homoscédasticité des données en entrée et le faible nombre d'exemples par classe.

L'ADL2DoC est également plus performante que les autres techniques à des rangs supérieurs : l'amélioration moyenne, calculée sur les sept premiers rangs par rapport aux *fisherfaces*, est d'environ 12%. On peut comparer ces performances à celles d'un tirage aléatoire, consistant à tirer au hasard et sans remise r individus parmi N. Sachant que la classe-cible Ω_j^* contient N_j^* exemple, on recherche la probabilité que, parmi les r boules tirées, au moins l'une d'entre elles appartienne à Ω_j^*. Cette probabilité p est égale à $p = 1 - q$, où q est la probabilité qu'aucune des r boules tirées n'appartienne à la classe Ω_j^*. Cette probabilité peut être modélisée par une loi hypergéométrique et l'on obtient :

$$p = 1 - \frac{C_{N_j^*}^0 C_{N-N_j^*}^r}{C_N^r} = 1 - \frac{(N - N_j^*)!(N-r)!}{N!(N - N_j^* - r)!} \tag{4.37}$$

Dans le contexte de cette expérimentation ($N = 321$ et $N_j^* = N_1 = \cdots = N_k = 3$), la probabilité p qu'au moins un des r visages tirés au hasard appartienne à la classe-cible est seulement de 0,93% au rang 1, de 4,6% au rang 5, de 9,1% au rang 10 et de 13,4% au rang 15 contre respectivement 72,4%, 86,4%, 95,3% et 97,2% pour l'ADL2DoC.

4.4.6.3 Expérimentations menées sur FERET et BioID

La troisième expérimentation (voir figure 4.9-c) est menée grâce aux bases FERET et BioID. On dispose de deux ensembles d'images de visages, dont on sait que chacun de ceux-ci correspond à une même personne. Ces images sont tirées de la base BioId et issues de séquences vidéo. On ne connaît aucune des deux personnes représentées, mais l'on désire simplement savoir si il s'agit de la même personne ou non. Pour cela, il nous faut construire en amont de la reconnaissance un modèle de visage suffisamment représentatif de l'ensemble des visages pour avoir une très bonne capacité de généralisation. On privilégiera donc une base d'apprentissage contenant suffisamment de données pour être représentative d'un maximum de variations possibles entre deux vues d'un même visage et entre des vues de deux visages différents. En l'occurrence, celle-ci est construite depuis la base FERET et contient 818 images de 152 personnes différentes. Le nombre d'images par personne est variable, mais toujours supérieur ou égal à quatre. Deux bases, issues de la base BioID, sont utilisées pour l'évaluation. Chacune d'entre elles contient 173 images de 18 personnes non enregistrées dans la base d'apprentissage. Pour une personne donnée, chacune de ces bases contient des images extraites d'une séquence différente. Les bases de connaissance et de test sont très dissimilaires en termes de conditions d'illumination, d'expression faciale et de pose de la tête (voir figure 4.9-c).

Dans un premier temps, chaque image issue de la base de test est comparée aux images de la base de connaissance suivant une distance Euclidienne au plus proche voisin. C'est ici l'ADL2DoL qui est la plus performante. La table 4.7 montre que les taux de reconnaissance de l'ACP2D, de l'ADL2DoL et des *fisherfaces* sont inférieurs à 65%, ce qui est relativement faible. On peut considérer que ces résultats mitigés proviennent des importantes dissimilarités entre les bases de

connaissance et de test. L'ADL2DoL permet cependant d'obtenir de bien meilleurs résultats que l'ACP2D et les *fisherfaces*, avec respectivement 14,5% et 16,2% d'amélioration.

Méthode	ADL2DoL	ADL2DoC	ACP2D	*fisherfaces*
Taux de reconnaissance	64,2%	61%	49,7%	48%

TAB. 4.7 – *Taux de reconnaissance comparés de l'ADL2DoC, de l'ADL2DoL, de l'ACP2D [YZFY04] et de l'algorithme des fisherfaces [BHK97] pour l'expérimentation utilisant PF01 et BioId (cf. figure 4.9-c).*

Dans un second temps, les bases de test et de connaissance sont comparées directement séquence à séquence, en appliquant la procédure de vote à la majorité suivante :

- pour chaque visage T d'une séquence issue de la base de test, on détermine son plus proche voisin X dans la base de connaissance ;
- on procède à un vote à la majorité : la séquence de la base de test est mise en correspondance avec la séquence de la base de connaissance dont sont le plus fréquemment issus les plus proches voisins.

Avec cette stratégie, l'ADL2DoL permet de bien classer onze séquences sur dix-huit tandis que l'ACP2D reconnaît au mieux huit séquences.

On peut déduire de ces résultats que l'ADL2Do semble avoir une meilleure capacité de généralisation à des personnes non enregistrées dans la base d'apprentissage que l'ACP2D et la méthode des *fisherfaces*.

4.4.7 Complémentarité de l'ADL2DoL et de l'ADL2DoC

Comme nous l'avons vu en section 4.4.6, suivant les bases de visages considérées, l'ADL2DoL peut se montrer plus performante que l'ADL2DoC, ou inversement. Tandis que l'ADL2DoL est plus efficace dans des conditions générales de prise de vue (bases ORL et BioId), l'ADL2DoC permet d'obtenir des résultats de classification sensiblement meilleurs que l'ADL2DoL en présence de variations importantes dans la pose et l'expression faciale (base PF01). Dans cette partie nous menons une analyse quantitative et qualitative plus poussée des performances des deux versions issues de l'ADL2Do, afin de mettre en évidence leur complémentarité. Nous utilisons pour cela la base de visages de Yale (*cf.* annexe A), qui contient 15 personnes et 11 vues par personne. Ces vues présentent des occultations partielles ainsi que des dissimilarités dans les conditions d'éclairage et les expressions faciales.

Dans la première expérimentation, une sous-base de Yale (*cf.* annexe A) contenant 15 personnes et dix vues par personne (toutes exceptée la vue « lumière droite ») est divisée aléatoirement en une base d'apprentissage contenant quatre vues par personne et une base de test contenant six vues par personne. Des extraits de la base de Yale sont donnés en figure 4.11. Pour favoriser l'homoscédasticité, on regroupe toutes les vues similaires dans la même base (apprentissage ou test). Cette opération est répétée cinq fois. Les matrices de confusion correspondantes sont présentées dans la table 4.8. La table 4.8-1 montre que, pour la première partition considérée, les performances des deux méthodes sont comparables : les taux de reconnaissance sont respectivement $\frac{53+10}{53+10+11+16} = 70\%$ et 71,1% pour l'ADL2DoL et l'ADL2DoC. Cependant, on peut noter que les résultats de classification sont très différents : 21 visages (23,3% de la base de test) sont

FIG. 4.11 – *Extraits de la base d'apprentissage et des sept bases de test extraites de Yale. Pour une personne donnée, si les vues de la base d'apprentissage sont non occultées alors la base « Occultation » contient une vue avec lunettes, et inversement.*

reconnus par une méthode seulement. On peut également noter que 82,2% $\gg max$(70%,71,1%) des visages sont reconnus par au moins l'une des deux méthodes. Les matrices de confusion (2-3) illustrent le fait que, généralement, l'ADL2DoL est plus performante que l'ADL2DoC. On peut remarquer que, dans ce cas, le taux de visages mal classés à la fois par les deux méthodes est faible (3,3% pour la partition (2) et 6,7% pour la (3)). Les matrices de confusion (4-5) montrent que, dans certains cas où le taux de visages mal classés par les deux méthodes est plus important ($\frac{16}{90}$ = 17,8% et 20% pour les partitions (4) et (5)), l'ADL2DoC est plus performante que l'ADL2DoL. Par conséquent, selon les caractéristiques des bases d'apprentissage et de test considérées, la méthode la plus performante n'est pas nécessairement la même.

$L \cap C_{(a)}$	$L \cap \overline{C}_{(b)}$
$\overline{L} \cap C_{(c)}$	$\overline{L} \cap \overline{C}_{(d)}$

(ref)

53	10
11	16

(1)

71	11
5	3

(2)

72	8
4	6

(3)

55	5
14	16

(4)

63	2
7	18

(5)

TAB. 4.8 – *Matrices de confusion de cinq partitions aléatoires de la base de Yale. Comme le montre la matrice (ref), l'élément noté (a) dans la matrice correspond au nombre de visages reconnus par les deux méthodes (ADL2DoL∩ADL2DoC, noté L∩C). L'élément (b) est le nombre de visages reconnus par l'ADL2DoL mais mal classés par l'ADL2DoC (L∩$\overline{C}$). L'élément (c) est le nombre de visages reconnus par l'ADL2DoC mais mal classés par l'ADL2DoL ($\overline{L} \cap C$). En (d), on trouve le nombre de visages mal classés par les deux méthodes ($\overline{L} \cap \overline{C}$). Les matrices de confusion (1-5) sont agencées de la même manière.*

La deuxième expérimentation fournit une analyse qualitative plus poussée. La base d'apprentissage, illustrée en figure 4.11, contient quatre vues pour chacune des 15 personnes, avec des changements dans les conditions d'illumination et les expressions faciales. Puis, sept bases de test sont construites (voir figure 4.11), à partir des vues restantes. La figure 4.12 donne une comparaison des performances de l'ADL2DoL et de l'ADL2DoC, sur ces sept bases de test. On constate que l'ADL2DoL est généralement plus performante que l'ADL2DoC. Cependant, dans certains cas, l'ADL2DoC est sensiblement meilleure que l'ADL2DoL, notamment quand la base de test contient des dissymétries selon l'axe vertical (par exemple pour les vues « Lumière Gauche » et « Lumière Droite »), ce qui semble logique au vu de l'interprétation géométrique donnée en section 4.4.2.4. L'ADL2DoC peut également fournir des résultats légèrement meilleurs si le changement d'expression faciale est très important, par exemple pour les vues « Surprise ».

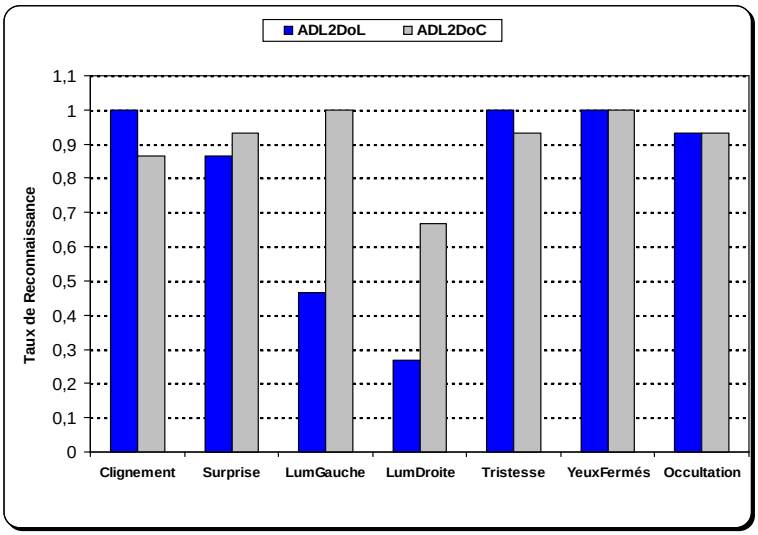


FIG. 4.12 – *Taux de reconnaissance comparés de l'ADL2DoL et de l'ADL2DoC, sur sept partitions de la base de Yale.*

4.4.8 Discussion

Cette section vise à mieux cerner le comportement des deux versions issues de l'ADL2Do, à souligner leurs avantages comme leurs inconvénients et à comparer leurs caractéristiques avec les principales approches de projection statistique.

Revenons sur l'analyse qualitative des résultats. L'ADL2DoL est en général plus performante que l'ADL2DoC. Elle nécessite également l'utilisation de moins de vecteurs propres. Cela peut être en partie expliqué par le fait que les images de visages sont des données symétriques selon un axe vertical et que, par conséquent, les lignes des visages contiennent plus d'information redondante que les colonnes. On comprend intuitivement que les hyperplans de séparation entre classes sont moins difficiles à estimer dans le cas de l'ADL2DoL que dans le cas de l'ADL2DoC. *A contrario*, dans le cas où les images-requêtes ont des valeurs de pixels asymétriques selon l'axe vertical (p. ex. pour les bases de test de Yale avec lumière de côté), l'ADL2DoC donne de meilleurs résultats car le changement d'illumination affecte seulement une partie des colonnes, alors qu'il affecte toutes les lignes.

Un avantage considérable de l'ADL2Do sur l'ADL unidimensionnelle (*cf.* chapitre 3.) est qu'elle ne souffre pas en général du problème de la singularité. Cela nous évite d'avoir à mettre en œuvre des solutions coûteuses, pouvant mener à la perte d'information discriminante, ou nécessitant l'ajustement difficile de paramètres supplémentaires. De plus, la taille réduite de la matrice de covariance permet une phase de construction plus rapide que pour les techniques 1D, ainsi qu'un coût de stockage réduit.

L'ADL2Do présente un désavantage majeur par rapport aux techniques 1D puisque, comme pour l'ACP2D, les signatures obtenues sont de taille supérieure. Prenons l'exemple de la base ORL. Avec cinq images, de taille 61×46 par personne, l'ADL2DoL et l'ADL2DoC nécessitent respectivement l'utilisation de $g = 5$ et $g = 10$ vecteurs propres, contre $g = 39$ pour les *fisherfaces*. Les signatures issues de l'ADL2DoL et de l'ADL2DoC sont donc respectivement des matrices de

Rang	**Performance**	**Coût de construction**	**Coût de classification**	**Coût de stockage**
1	ADL2Do, ACP+ADL$_0$	ADL2Do, ACP2D	ACP+ADL$_0$	ACP2D, ADL2Do
2	ACP2D	ADLD	*fisherfaces*, ADLD	ACP+ADL$_0$
3	*fisherfaces*	*eigenfaces*	*eigenfaces*	*eigenfaces*, ADLD
4	ADLD	*fisherfaces*	ADL2Do, ACP2D	*fisherfaces*
5	*eigenfaces*	ACP+ADL$_0$		

TAB. 4.9 – *Classement des principales méthodes présentées dans cette section, de la plus efficace à la moins efficace, en fonction de quatre critères. Les méthodes testées sont : l'ADL2Do (ADL2DoL et ADL2DoC) l'ACP2D [YZFY04], l'ADL dans le noyau de Huang et al. [HLLM02] (ACP+ADL$_0$), l'ADL Directe (ADLD) [YY01], ainsi que les techniques des eigenfaces [TP91] et des fisherfaces [BHK97].*

taille 61×5 (soit 305 coefficients) et 46×10 (460 coefficients). Les signatures fournies par les *fisherfaces* sont des vecteurs constitués de seulement 39 éléments. La phase de classification est donc plus coûteuse en termes de temps de calcul pour l'ADL2Do que pour les *fisherfaces*. Ceci constitue un désavantage pouvant être considéré comme pénalisant pour l'ADL2Do, car la phase de comparaison de signatures est souvent une étape menée *en-ligne*.

La table 4.9 donne un classement au vu de quatre critères de l'ADL2Do (meilleure des deux versions), de l'ACP2D et des principales techniques d'ADL utilisant une modélisation 1D des données (*cf.* section 3.3.3). Le premier critère étudié est la performance des algorithmes. Le classement est effectué au vu des taux de reconnaissance obtenus sur la base ORL (*cf.* table 4.5) et des enseignements tirés des chapitres 2. et 3. Les deuxième et troisième critères correspondent respectivement au coût de la construction du modèle et de la classification, en termes de nombres d'opérations du système. Le quatrième critère mesure le coût de stockage du modèle.

On peut noter qu'en cette année 2005 les techniques Bidimensionnelles orientées ont fait l'objet d'un certain engouement. En Juillet 2005, Xiong *et al.* ont proposé dans [XSA05] une méthode baptisée Two-dimensional Fisher Linear Discriminant Analysis *2DFLD*, strictement équivalente à l'ADL2DoL (que nous avions précédemment introduite en Septembre 2004 à la conférence internationale ICCVG [VGJ04]). En 2005, Zhou a proposé dans un article actuellement soumis au journal *IEEE Transactions on Pattern Analysis and Machine Intelligence* une technique généralisant l'ACP2D ainsi que les deux versions de l'ADL2Do à des projections non linéaires par l'utilisation d'une fonction de noyau [Zho05]. Les résultats expérimentaux montrent une amélioration des taux de reconnaissance en présence d'importantes variations d'illumination. Notons que les effets de l'illumination sur la technique d'ADL2Do seront évalués en section 5.2.5.3.

4.5 Conclusion

La plupart des méthodes globales de reconnaissance de visages basées sur la projection statistique consistent à mettre en œuvre une ou plusieurs techniques d'analyse de données sur les images de visages, modélisées par le biais de vecteurs. Comme nous l'avons montré au chapitre 3.

ce mode de représentation (1D) des données en entrée engendre un certain nombre de difficultés pour la mise en œuvre directe de l'ADL (problème de la singularité notamment).

Par l'étude de la technique d'ACP2D introduite par Yang *et al.* [YZFY04] nous avons dans un premier temps montré les avantages de la modélisation 2Do des données en termes de performance, de coût de construction du modèle et de tolérance à quelques-unes des principales sources de variation des images de visages. Or, nous avions souligné aux chapitres 2. et 3. que l'ADL est plus généralement plus performante pour la tâche d'identification que l'ACP.

C'est pourquoi nous avons introduit dans un second temps une nouvelle technique de classification, à savoir l'Analyse Discriminante Linéaire 2D orientée (ADL2Do), qui allie les avantages de la représentation 2Do des données et le pouvoir discriminant de l'ADL. Cette technique, basée sur une généralisation du critère de Fisher à des matrices représentatives des dispersions intra- et inter-classe (orientées), évite implicitement le problème de la singularité récurrent dans le cas 1D. Elle permet d'obtenir des taux de reconnaissance comparables aux meilleures techniques de projection statistique, dans un contexte d'identification en monde fermé. Nous avons de plus montré sa tolérance à des variations de pose et d'expression faciale. L'ADL2Do se décline en deux versions : l'une orientée en lignes, et l'autre en colonnes, dont nous avons montré l'efficacité et la complémentarité en termes de résultats de classification. Ce constat ouvre la voie à différents modes de fusion et/ou de combinaison de ces deux techniques dans le but de construire un classifieur réellement bidimensionnel, c'est-à-dire basé conjointement sur les modélisations orientées en lignes et en colonnes des visages. La méthode ainsi construite devra pallier le principal inconvénient de la technique de l'ADL2Do, qui réside dans une classification plus coûteuse que pour les méthodes unidimensionnelles. Cette étude fait l'objet du chapitre suivant.

Chapitre 5

Une nouvelle approche Discriminante Bidimensionnelle en monde fermé ou ouvert

5.1 Introduction

Dans le chapitre précédent, nous avons introduit une nouvelle technique d'extraction de signatures, baptisée ADL2Do. Cette méthode se décline en deux versions, à savoir l'ADL2DoL et l'ADL2DoC, dont nous avons montré la complémentarité des résultats de classification. Combiner efficacement ces deux techniques peut donc engendrer une approche plus performante.

La section 5.2 de ce chapitre vise à introduire une telle méthode, que nous baptisons *Analyse Discriminante Bilinéaire* (ADB). Cette approche originale est considérée comme bidimensionnelle, car tirant avantage à la fois de la modélisation orientée en lignes et en colonnes. Elle corrige le principal désavantage des méthodes d'ADL2Do, à savoir une classification relativement coûteuse par rapport aux autres techniques de projection statistique. Nous montrerons que l'ADB est plus performante que l'ADL2Do et que les principales techniques de projection statistique de l'état de l'art. Afin d'améliorer la tolérance de l'ADB à des changements d'expression faciale, nous détaillons son utilisation dans le cadre d'une technique hybride de combinaison modulaire d'experts. La nouvelle technique ainsi définie, nommée ADB Modulaire (ADBM), fait l'objet de la section 5.3. L'ADB, tout comme l'ADBM, sont des techniques d'extraction de signatures. Les signatures obtenues sont habituellement classées à l'aide d'une mesure de dissimilarité au plus proche voisin. Ce mode de mise en correspondance souffre de deux désavantages principaux. Premièrement, la règle au plus proche voisin est très coûteuse. Deuxièmement, on ne peut pas directement l'utiliser dans le contexte d'une application en monde ouvert (*cf.* section 1.3). Pour cela, il nous faudrait définir une valeur maximale de cette mesure, au-delà de laquelle le visage-requête n'appartient pas à la base d'apprentissage. Étant donné que ces mesures de dissimilarité ne sont généralement pas normalisées (leur amplitude diffère d'une base d'apprentissage à l'autre), le choix du seuil est un problème difficile. Afin de pallier ces inconvénients, nous introduisons en section 5.4 l'utilisation d'un Réseau de Fonctions à Base Radiale Normalisé pour une classification plus rapide et efficace, notamment en monde ouvert.

5.2 L'analyse Discriminante Bilinéaire

5.2.1 Introduction

Divers schémas de combinaison des deux approches de l'ADL2Do ont été évalués. Nous présentons dans cette section la technique qui s'est révélée la plus efficace en termes de performances, que nous baptisons *Analyse Discriminante Bilinéaire* (ADB). Nous utilisons le qualificatif *bilinéaire* car la projection linéaire utilisée dans le contexte de l'Analyse Discriminante Linéaire est remplacée ici par une projection bilinéaire. Au lieu de rechercher l'espace de projection séparant au mieux par projection linéaire les différentes classes, on cherche le couple de matrices de projection qui, par projection bilinéaire, permet de classer au mieux les données.

5.2.2 Extraction de signatures

Les notations sont les mêmes qu'au chapitre 4. Considérons deux matrices de projection $Q \in \mathbb{R}^{h \times g}$ et $P \in \mathbb{R}^{w \times g}$, où g est un paramètre du modèle. Définissons la signature d'une image X_l comme étant sa projection bilinéaire sur le couple de matrices (Q,P) :

$$X_l^{(Q,P)} = Q^T X_l P \tag{5.1}$$

Notre but est de déterminer le couple de matrices de projection (Q,P) optimal, maximisant la séparation entre signatures provenant de différentes classes tout en minimisant la séparation entre signatures d'une même classe, c.-à-d. maximisant le critère de Fisher (*cf.* section 3.2.3) suivant :

$$J(Q,P) \quad = \quad \frac{|S_b^{(Q,P)}|}{|S_w^{(Q,P)}|} \tag{5.2}$$

où :

- $S_w^{Q,P}$ et $S_b^{Q,P}$ sont respectivement les matrices de covariance intra-classe et inter-classe évaluées à partir de l'ensemble $(X_l^{Q,P})_{l \in \{1,\ldots,N\}}$ des images de la base d'apprentissage projetées :

$$S_w^{(Q,P)} = \frac{1}{N} \sum_{j=1}^{k} \sum_{X_l \in \Omega_j} (X_l^{(Q,P)} - \overline{X_j^{(Q,P)}})^T (X_l^{(Q,P)} - \overline{X_j^{(Q,P)}}) \tag{5.3}$$

$$S_b^{(Q,P)} = \frac{1}{N} \sum_{j=1}^{k} N_j (\overline{X_j^{(Q,P)}} - \overline{X^{(Q,P)}})^T (\overline{X_j^{(Q,P)}} - \overline{X^{(Q,P)}}) \tag{5.4}$$

- $\overline{X_j^{(Q,P)}} = \frac{1}{N_j} \sum_{X_l \in \Omega_j} X_l^{(Q,P)}$ est la moyenne de tous les exemples projetés correspondant à la classe j ;
- $\overline{X^{(Q,P)}} = \frac{1}{N} \sum_{j=1}^{k} N_j \overline{X_j^{(Q,P)}}$ est la moyenne de tous les exemples de la base d'apprentissage Ω, projetés sur (Q,P).

Développons l'expression (5.2) :

$$J(Q,P) \quad = \quad \frac{\frac{1}{N} | \sum_{j=1}^{k} N_j P^T (\overline{X_j} - \overline{X})^T Q Q^T (\overline{X_j} - \overline{X}) P |}{\frac{1}{N} | \sum_{j=1}^{k} \sum_{X_l \in \Omega_j} P^T (X_l - \overline{X_j})^T Q Q^T (X_l - \overline{X_j}) P |} \tag{5.5}$$

Cette fonction objectif (5.5) est biquadratique (selon le couple de matrices (*Q*,*P*)), et n'a par conséquent pas de solution analytique. C'est pourquoi nous proposons une procédure itérative, que nous nommons *Analyse Discriminante Bilinéaire*, et qui a fait l'objet d'une publication dans [VGJ05a].

Quelle que soit la matrice $Q \in \mathbb{R}^{h\times g}$ fixée, la matrice P maximisant le critère $J(Q,P)$ donné en équation (5.5) maximise le critère suivant :

$$\begin{aligned} J_Q(P) &= \frac{|P^T \left[\sum_{j=1}^{k} N_j(\overline{X_j^Q} - \overline{X^Q})^T(\overline{X_j^Q} - \overline{X^Q})\right] P|}{|P^T \left[\sum_{j=1}^{k}\sum_{X_l\in\Omega_j} (X_l^Q - \overline{X_j^Q})^T(X_l^Q - \overline{X_j^Q})\right] P|} \\ J_Q(P) &= \frac{|P^T S_b^Q P|}{|P^T S_w^Q P|} \end{aligned} \tag{5.6}$$

où :

- S_w^Q et S_b^Q sont respectivement les matrices de dispersion intra-classe et inter-classe généralisées des visages de Ω projetés sur Q selon l'équation (4.17) ($\forall l = 1, \ldots, N$, $X_l^Q = Q^T X_l$) :

$$S_w^Q = \frac{1}{N}\sum_{j=1}^{k}\sum_{X_l\in\Omega_j} (X_l^Q - \overline{X_j^Q})^T(X_l^Q - \overline{X_j^Q}) \tag{5.7}$$

$$S_b^Q = \frac{1}{N}\sum_{j=1}^{k} N_j(\overline{X_j^Q} - \overline{X^Q})^T(\overline{X_j^Q} - \overline{X^Q}) \tag{5.8}$$

- $\overline{X_j^Q} = \frac{1}{N_j}\sum_{X_l\in\Omega_j} X_l^Q$ est la moyenne de tous les exemples de la classe j, projetés sur Q selon (4.17) ;
- $\overline{X^Q} = \frac{1}{N}\sum_{j=1}^{k} N_j\overline{X_j^Q}$ est la moyenne de tous les exemples projetés de la base d'apprentissage Ω.

Par conséquent, pour toute matrice $Q \in \mathbb{R}^{h\times g}$ fixée, la matrice P maximisant le critère J_Q donné en équation (5.6) est la matrice de taille $w \times g$ dont les colonnes sont les vecteurs propres de la matrice $S_w^{Q^{-1}} S_b^Q$, associés aux plus grandes valeurs propres. Remarquons que, si Q est la matrice identité de taille $h \times h$, alors P est la matrice de projection de l'ADL2DoL, présentée en section 4.4.2.1.

Notons $A = P^T(\overline{X_j} - \overline{X})^T Q$, matrice carrée de taille $g \times g$. Etant donné que, pour chaque matrice carrée A, $|A^T A| = |AA^T|$, la fonction objectif (5.5) peut être réécrite :

$$J(Q,P) = \frac{|\sum_{j=1}^{k} N_j Q^T(\overline{X_j} - \overline{X})PP^T(\overline{X_j} - \overline{X})^T Q|}{|\sum_{j=1}^{k}\sum_{X_l\in\Omega_j} Q^T(X_l - \overline{X_j})PP^T(X_l - \overline{X_j})^T Q|} \tag{5.9}$$

Donc, pour toute matrice $P \in \mathbb{R}^{w\times g}$ fixée, la matrice Q maximisant le critère (5.5) est la matrice de taille $h \times g$ maximisant le critère suivant :

$$J_P(Q) = \frac{|Q^T \Sigma_b^P Q|}{|Q^T \Sigma_w^P Q|} \tag{5.10}$$

où Σ_w^P et Σ_b^P sont les matrices de dispersion intra-classe et inter-classe généralisées associées aux transposées des images de Ω, projetées sur P selon l'équation (4.1) ($\forall l = 1, \ldots, N$, $X_l^P = X_l P$) :

$$\Sigma_w^P = \frac{1}{N} \sum_{j=1}^{k} \sum_{X_l \in \Omega_j} (X_l^P - \overline{X_j^P})(X_l^P - \overline{X_j^P})^T \tag{5.11}$$

$$\Sigma_b^P = \frac{1}{N} \sum_{j=1}^{k} N_j (\overline{X_j^P} - \overline{X^P})^T (\overline{X_j^P} - \overline{X^P}) \tag{5.12}$$

où $\overline{X_j^P} = \frac{1}{N_j} \sum_{X_l \in \Omega_j} X_l^P$ est la moyenne de tous les exemples projetés X_l^P de la classe Ω_j et $\overline{X^P} = \frac{1}{N} \sum_{j=1}^{k} N_j \overline{X_j^P}$ est la moyenne de tous les exemples projetés de la base d'apprentissage Ω.

Par conséquent, les colonnes de Q sont les g vecteurs propres de $(\Sigma_w^P)^{-1}\Sigma_b^P$ associées aux plus grandes valeurs propres. Notons que, dans le cas où P est la matrice identité de taille $w \times w$, la matrice Q est la matrice de projection obtenue par l'ADL2DoC (*cf.* section 4.4.2.2).

5.2.2.1 Algorithmes proposés

La technique d'ADB proposée consiste en un processus itératif mettant en œuvre alternativement une ADL2DoL et une ADL2DoC sur les images. Dans un premier temps, une ADL2DoL est appliquée sur les images initiales de manière à obtenir une matrice de projection optimale en lignes P. Dans un deuxième temps, on projette les images initiales sur cet espace de projection. Puis, ces données projetées sont utilisées comme base d'apprentissage d'un modèle d'ADL2DoC ; on obtient la matrice de projection Q. Les images initiales sont alors projetées sur Q ; à partir de ces données projetées on construit un modèle d'ADL2DoL, et ainsi de suite. Notons que l'ordre de mise en œuvre de l'ADL2DoL et de l'ADL2DoC peut être inversé. L'algorithme résultant appliquerait d'abord une ADL2DoC sur les images initiales.

Nous proposons deux algorithmes itératifs pour la mise en œuvre de l'ADB. Ces algorithmes, appelés Algorithme 1 (ADB1) et Algorithme 2 (ADB2), sont donnés ci-après. Le premier est construit à partir d'un nombre g de vecteurs propres fixé, qui peut être déterminé à l'aide d'un échantillon de validation, comme détaillé en section 4.4.4. Le second algorithme permet de sélectionner automatiquement le nombre g de vecteurs propres optimal, par suppression séquentielle des vecteurs les moins discriminants. Ces algorithmes ont été détaillés dans [VGJ05d].

Algorithme 1 ADB1

Entrées :
X : ensemble des images de Ω ;
Id : ensemble des identités associées aux images de X ;
g : nombre de vecteurs de projection retenus dans chaque matrice de projection ;
Sorties :
P : matrice de projection à droite dans (5.1) ;
Q : matrice de projection à gauche dans (5.1) ;
Initialisation :
$Q_0 \leftarrow I_h$ avec I_h matrice identité de taille $h \times h$;
Début :
Pour $t = 1$ à T **faire**
pour tout $l \in \{1, \ldots, N\}$, calculer $X_l^{Q_{t-1}} = Q_{t-1}^T X_l$;
calculer $S_w^{Q_{t-1}}$, $S_b^{Q_{t-1}}$ et $(S_w^{Q_{t-1}})^{-1} S_b^{Q_{t-1}}$;
calculer le système propre (V_i, λ_i) de $(S_w^{Q_{t-1}})^{-1} S_b^{Q_{t-1}}$;
ranger les vecteurs propres V_i en ordre décroissant de leur valeur propre λ_i associée ;
construire la matrice $P_t = [V_1, \ldots, V_g]$;
pour tout $l \in \{1, \ldots, N\}$, calculer $X_l^{P_t} = X_l P_t$;
calculer $\Sigma_w^{P_t}$, $\Sigma_b^{P_t}$ et $(\Sigma_w^{P_t})^{-1} \Sigma_b^{P_t}$;
calculer le système propre (V_i', λ_i') de $(\Sigma_w^{P_t})^{-1} \Sigma_b^{P_t}$;
ranger les vecteurs propres V_i' en ordre décroissant de leur valeur propre λ_i' associée ;
construire la matrice $Q_t = \left[V_1', \ldots, V_g'\right]$;
fin Pour
$P \leftarrow P_T$;
$Q \leftarrow Q_T$;
Fin

Le nombre g de vecteurs propres à utiliser peut être déterminé à l'aide d'une base de validation indépendante de la base d'apprentissage (*cf.* section 4.4.4).

Nos résultats expérimentaux ont montré qu'après $T = 1$ itération seulement les taux de reconnaissance sont satisfaisants, et relativement stables vis-à-vis du paramètre g (des changements raisonnables dans la valeur de g n'ont pas d'impact important sur les performances). Par conséquent, dans la suite, nous utiliserons cet algorithme avec une seule itération, ce qui nous évite d'avoir à déterminer le τ optimal, tout en garantissant de bonnes performances.

Algorithme 2 ADB2

Entrées :
X : ensemble des images de Ω ;
Id : ensemble des identités associées aux images de X ;
$init_g$: nombre de vecteurs de projection à l'initialisation ;
Sorties :
P : matrice de projection à droite dans (5.1) ;
Q : matrice de projection à gauche dans (5.1) ;
Initialisation :
$t \leftarrow 0$;
$Q_0 \leftarrow I_h$ avec I_h matrice identité de taille $h \times h$;
$g_0 \leftarrow init_g$, $p_0 \leftarrow 1$ et $p_1 \leftarrow 1$;
Début :
Tant que $p_t > 0{,}05$ **faire**
$t \leftarrow t + 1$;
$g_t \leftarrow g_{t-1} - 1$;
pour tout $l \in \{1, \ldots, N\}$, calculer $X_l^{Q_{t-1}} = Q_{t-1}^T X_l$;
calculer $S_w^{Q_{t-1}}$, $S_b^{Q_{t-1}}$ et $(S_w^{Q_{t-1}})^{-1} S_b^{Q_{t-1}}$;
calculer le système propre (V_i, λ_i) de $(S_w^{Q_{t-1}})^{-1} S_b^{Q_{t-1}}$;
ranger les vecteurs propres V_i en ordre décroissant de leur valeur propre λ_i associée ;
construire la matrice $P_t = [V_1, \ldots, V_g]$;
pour tout $l \in \{1, \ldots, N\}$, calculer $X_l^{P_t} = X_l P_t$;
calculer $\Sigma_w^{P_t}$, $\Sigma_b^{P_t}$ et $(\Sigma_w^{P_t})^{-1} \Sigma_b^{P_t}$;
calculer le système propre (V_i', λ_i') de $(\Sigma_w^{P_t})^{-1} \Sigma_b^{P_t}$;
ranger les vecteurs propres V_i' en ordre décroissant de leur valeur propre λ_i' associée ;
construire la matrice $Q_t = \left[V_1', \ldots, V_g'\right]$;
Si $t > 1$ **alors**
$p_t =$`Wilks-Lambda`$(\left[V_{g_t+1}', \ldots, V_{g_0}'\right])$;
fin Si
fin Pour
$P \leftarrow P_{t-1}$;
$Q \leftarrow Q_{t-1}$;
Fin

La procédure `Wilks-Lambda` désigne le test de Wilks-Lambda, détaillé en section 4.4.4.1 et appliqué sur les $g_0 - g_t$ derniers vecteurs propres de $(\Sigma_w^{P_t})^{-1} \Sigma_b^{P_t}$. Si la p-valeur p_t (*cf.* note de bas de page n° 14 en p. 108) est supérieure à $\alpha = 5\%$, alors on ne peut pas rejeter l'hypothèse H_0. Les derniers vecteurs propres (non retenus dans la matrice de projection Q_t) sont donc considérés comme non porteurs d'information discriminante, et l'algorithme peut continuer. À l'inverse, si $p_t < 5\%$, alors on rejette H_0 et les derniers vecteurs propres sont discriminants. L'algorithme s'arrête et retourne le dernier couple de matrices de projection n'engendrant pas de perte d'information discriminante, à savoir (P_{t-1}, Q_{t-1}).

Le choix du paramètre d'initialisation $init_g$ est important, car il détermine à la fois la complexité de l'algorithme en termes de nombre d'itérations nécessaires à la convergence et le nombre de degrés de liberté d du test de Wilks-Lambda (*cf.* section 4.4.4.1). Nos expérimentations montrent qu'une valeur de $init_g = k$ garantit de très bons résultats dans tous les cas. Cependant, si le nombre k de classes est important, une initialisation à $init_g = k$ peut être inutilement coû-

teuse en termes de nombres d'itérations. Ces expériences ont montré que, pour les tailles de base envisagées dans le contexte de cette thèse (inférieures à 200, *cf.* section 1.3) et avec une résolution d'images voisine de 75×65 (*cf.* section 5.2.4.1), on peut imposer sans baisse de performance la contrainte supplémentaire $init_g \leq 40$:

$$init_g = \min(k, 40)$$

Les deux algorithmes ci-dessus peuvent directement être modifiés de manière à inverser l'ordre d'application de l'ADL2DoL et de l'ADL2DoC. Nos résultats expérimentaux ont montré que cet ordre n'a pas d'impact significatif sur les performances du système.

5.2.2.2 Interprétation géométrique

L'ADB peut être vue comme un algorithme d'extraction de sous-espace discriminant en deux temps : dans un premier temps, on retient l'information la plus discriminante pour les lignes des images de visages. Ce faisant, on construit l'espace de projection discriminant, au sens des lignes des images. Dans un second temps, on projette les images de la base d'apprentissage dans l'espace de projection ainsi construit. On obtient une base de signatures de visages dont les lignes sont « débruitées » (si l'on considère que le bruit est constitué de l'information non discriminante). On construit à partir de cette base de visages projetés un espace de projection discriminant, au sens des colonnes des visages débruités en lignes. Ainsi, l'ADB peut être vue comme une technique permettant de retirer itérativement le bruit des lignes, et des colonnes de l'image.

5.2.3 Classification

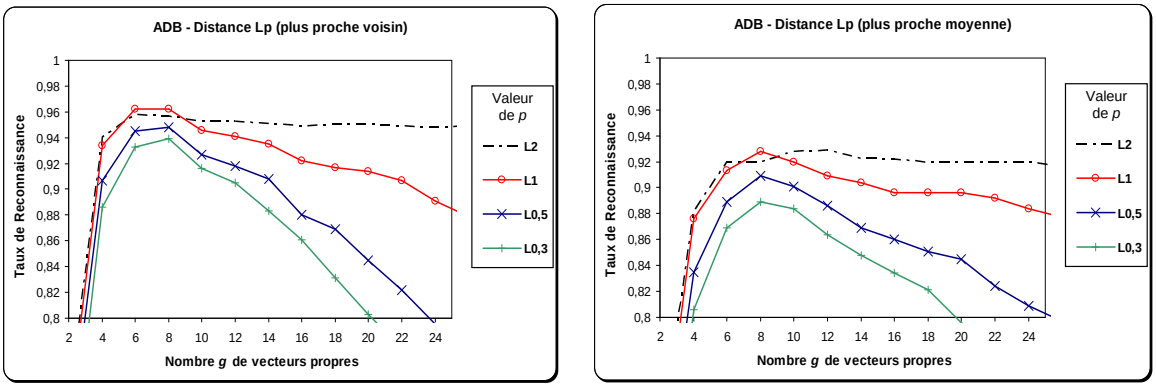
De la même manière qu'en section 4.4.3, nous cherchons à déterminer quelle est la mesure de dissimilarité la mieux adaptée à la classification des signatures issues de l'ADB. Les mesures de dissimilarité évaluées sont les distances et mesures fractionnaires de Minkowski suivantes :

$$D_{L_p}^{(Q,P)}(X^{(Q,P)}, Y^{(Q,P)}) = \sqrt[p]{\sum_{i=1}^{g}\sum_{m=1}^{g} |X^{(P,Q)}(i,m) - Y^{(P,Q)}(i,m)|^p} \qquad (5.13)$$

où $X^{(P,Q)}(i,m)$ est l'élément correspondant à la $i^{\text{ème}}$ ligne et $m^{\text{ème}}$ colonne de la matrice $X^{(P,Q)}$, et $p \in]0,1] \cup \{2\}$. Si $p = 1$ ou $p = 2$, l'équation (5.13) définit les distances de Minkowski usuelles, tandis que pour $p \in]0,1[$ il s'agit des mesures de dissimilarité fractionnaires (*cf.* annexe D).

On évalue les stratégies d'affectation au plus proche voisin et à la plus proche moyenne. Le protocole expérimental retenu est le même que celui utilisé en section 4.4.3. Les taux de reconnaissance moyens sur les cinq partitions d'ORL, au plus proche voisin et à la plus proche moyenne, sont donnés respectivement en figure 5.1-a et 5.1-b. Dans un souci de simplicité des notations, les distances $D_{L_p}^{(Q,P)}$ seront désignées dans la suite par D_{L_p}.

Les enseignements que l'on peut tirer de ce graphe sont globalement les mêmes que pour l'ADL2Do (*cf.* section 4.4.3.2). La règle d'affectation au plus proche voisin est plus efficace que la plus proche moyenne, et ceci quelle que soit la mesure de dissimilarité considérée. De plus, la classification par distance D_{L_2} est moins dépendante du nombre g de vecteurs propres retenus que les mesures D_{L_p} où $p < 2$. Dans la suite, nous utiliserons donc la distance Euclidienne D_{L_2} au plus proche voisin.



(a) **(b)**

FIG. 5.1 – *Impact du paramètre p de la mesure de dissimilarité D_{L_p} sur les taux de reconnaissance moyens calculés sur cinq partitions de la base ORL. Les techniques évaluées en (a) et (b) sont respectivement l'ADB (Algorithme 1) avec une stratégie d'affectation au plus proche voisin, et à la plus proche moyenne.*

5.2.4 Impact de différents facteurs sur les performances du système

Tout comme l'ADL2Do (*cf.* section 4.4.5), les performances de l'ADB peuvent être influencées par le choix de la résolution des images, le nombre de personnes enregistrées dans la base, et le nombre d'images par classe. Cette section vise à étudier l'impact de ces différents facteurs sur les performances du système, afin de déterminer leurs valeurs optimales.

5.2.4.1 Impact de la résolution des images

Le protocole expérimental retenu pour cette évaluation est le même que pour l'ADL2Do, en section 4.4.5.1. La figure 5.2 montre que, tout comme pour l'ADL2Do, la résolution des images semble ne pas affecter de manière importante les taux de reconnaissance, jusqu'à un certain point. Choisir une résolution d'images de 75×65 pixels semble constituer un bon compromis

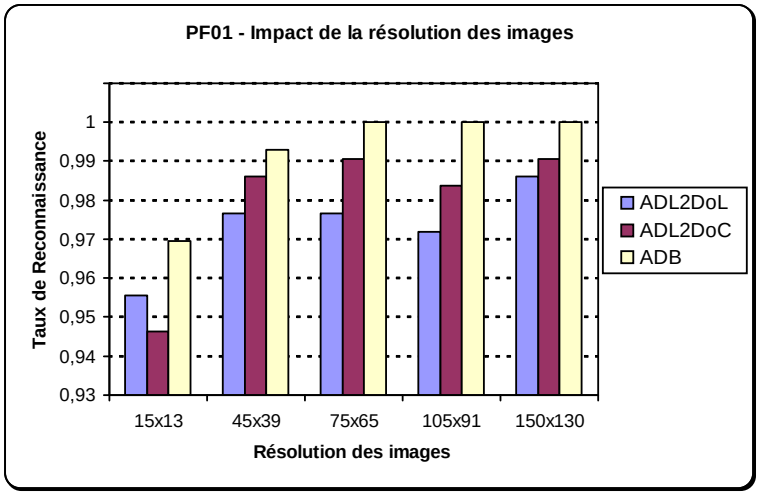
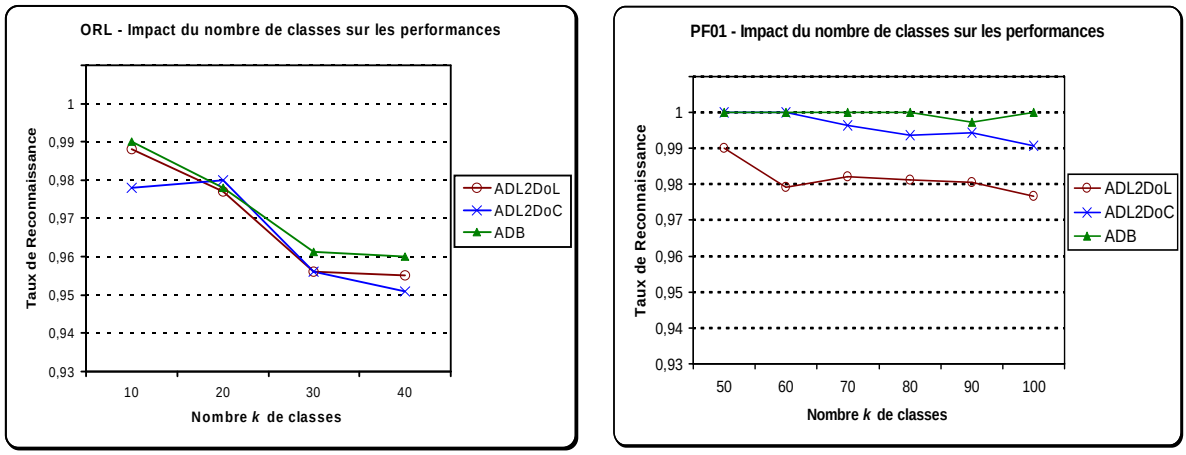


FIG. 5.2 – *Taux de reconnaissance comparés de l'ADB, de l'ADL2DoL et de l'ADL2DoC sur un sous-ensemble de la base PF01, à différentes résolutions.*

entre tailles des signatures et performances. C'est donc une résolution voisine de celle-ci (ou légèrement inférieure pour des bases de taille plus faible) que nous retiendrons dans la suite de nos expérimentations.

5.2.4.2 Impact du nombre de personnes enregistrées

Dans cette section, nous cherchons à caractériser l'impact du nombre de personnes enregistrées sur les performances de l'ADB. Le protocole expérimental retenu est le même que pour l'ADL2Do (voir section 4.4.5.2). Les graphes 5.3-a et 5.3-b montrent que l'ADB semble être moins influencée par le nombre k de classes que les deux versions de l'ADL2Do ; néanmoins sur ORL on note une baisse des performances lorsque le nombre de classes augmente. En ce qui concerne PF01, le taux de reconnaissance de l'ADB est de 100%, et ceci quel que soit le nombre de classes considérées.



(a) **(b)**

FIG. 5.3 – *Impact du nombre de classes de la base d'apprentissage sur les taux de reconnaissance de l'ADB, de l'ADL2DoL et de l'ADL2DoC évalués (a) sur la base ORL et (b) sur le sous-ensemble de la base PF01 décrit en section 4.4.5.1.*

5.2.4.3 Impact du nombre d'exemples par classe

Le protocole expérimental mis en œuvre est le même qu'en section 4.4.5.3. La figure 5.4 montre que, comme pour toutes les techniques de projection statistique, les performances de l'ADB sont très influencées par le nombre n d'exemples par classe. Néanmoins, sur la base considérée, les taux de reconnaissance atteignent 100% avec 4 images par classe. D'autres expérimentations, menées sur des bases contenant plus de sources de variations, montrent toutes un accroissement des taux de reconnaissance avec le nombre d'exemples par classe. Le point d'inflexion de la courbe, passé lequel les taux de reconnaissance augmentent moins fortement, est généralement situé en $n = 5$.

5.2.5 Évaluation des performances et comparaison aux techniques usuelles de projection statistique

Dans cette section, nous présentons les résultats d'expérimentations menées sur les bases ORL, FERET et AR (*cf.* annexe A). Nous utilisons la base ORL pour vérifier l'efficacité de l'ADB en tant que mode de combinaison de l'ADL2DoL et de l'ADL2DoC, pour comparer les

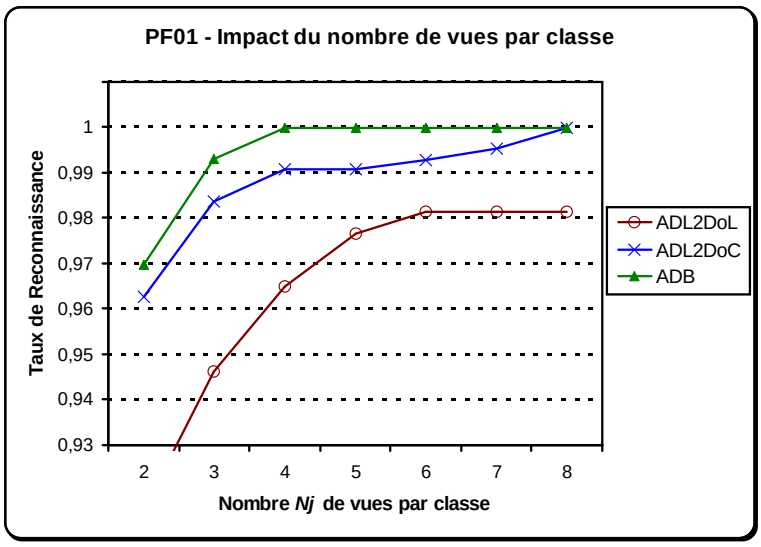


FIG. 5.4 – *Taux de reconnaissance comparés de l'ADB, de l'ADL2DoL et de l'ADL2DoC sur un sous-ensemble de la base PF01, lorsque le nombre N_j de vues par classe varie.*

performances des deux algorithmes ADB1 et ADB2 et situer les performances de l'ADB par rapport aux principales approches de l'état de l'art. L'expérience menée sur la base FERET permet de comparer les pouvoirs de généralisation (à des personnes non enregistrées dans la base d'apprentissage) de l'ADB à l'ADL2Do et à l'ACP2D. Les expérimentations menées sur la base AR visent à comparer les performances de ces trois méthodes en présence de variations d'éclairage.

5.2.5.1 Expérimentation menée sur la base ORL

Le protocole expérimental utilisé est le même qu'en section 4.4.6.1. Rappelons que la base ORL utilisée contient des variations limitées de la pose de la tête, de l'expression faciale et des conditions d'illumination. Elle est divisée aléatoirement en deux sous-bases, chacune contenant cinq images par personne pour les 40 personnes enregistrées. Pour chaque partition aléatoire, on calcule les taux de reconnaissance en utilisant la distance D_{L_2} au plus proche voisin (*cf.* équations (5.13) et 4.26).

La figure 5.5 donne les taux de reconnaissance comparés de l'ADB (Algorithmes 1 et 2), de l'ADL2DoL, de l'ADL2DoC et de l'ACP2D, pour l'une de ces partitions aléatoires, en fonction du nombre g de vecteurs propres retenus. Tandis que, pour construire ce graphe, il a été nécessaire de relancer plusieurs fois l'ADB1, les valeurs données pour l'ADB2 correspondent aux différentes itérations (à lire pour g décroissant, de la droite vers la gauche) d'une même occurrence de l'algorithme, avec $init_g = 40$. Ce graphique montre la supériorité de l'ADB sur les autres méthodes pour la partition considérée, ceci indépendamment de l'algorithme (1 ou 2) utilisé. L'ADB2 présente le désavantage de nécessiter l'utilisation de plus de vecteurs propres pour fournir une performance optimale. Les signatures fournies par l'ADB2 sont donc de taille plus importante que celles issues de l'ADB1. De plus, l'ADB2 nécessite plus d'itérations (une quinzaine contre une seule pour l'ADB1) et est donc plus coûteuse en termes de construction du modèle. C'est pourquoi, dans la suite, nous utiliserons préférentiellement l'algorithme 1. Remarquons également que l'ADL2DoL est plus performante que l'ACP2D, elle-même dépassant l'ADL2DoC, pour la partition considérée.

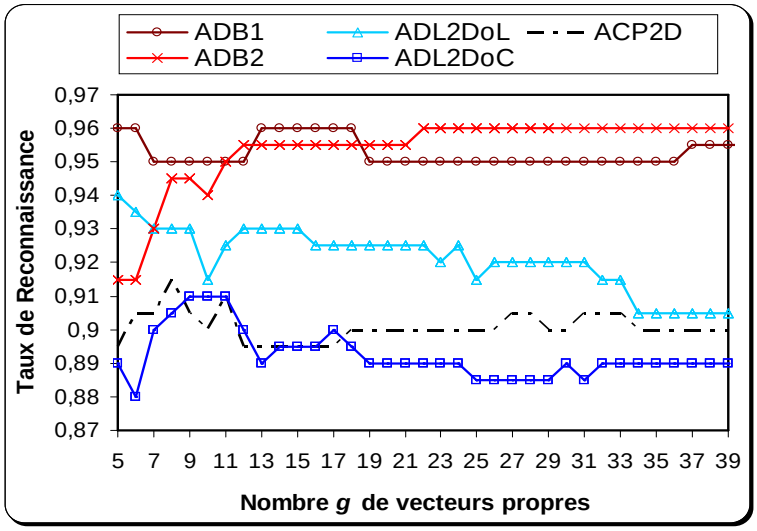


FIG. 5.5 – *Comparaison des taux de reconnaissance des six méthodes calculés sur la partition (1), en fonction du nombre g de vecteurs propres retenus.*

Intéressons-nous maintenant à l'ADB en tant que mode de combinaison des deux versions issues de l'ADL2Do. La table de contingence, dont les effectifs sont sommés sur les dix partitions testées (soit un nombre total de $10 \cdot 200 = 2000$ visages à reconnaître), est donnée en table 5.1. Le symbole logique « $\rceil$ » signifie « non », c.-à-d. que le nombre en deuxième ligne et première colonne du tableau, à savoir 40, est le nombre total (sur 2000) de visages reconnus par l'ADL2DoL, mais pas par l'ADL2DoC. Le symbole logique $\cap$ signifie « et » : le nombre en deuxième ligne et deuxième colonne, à savoir 24, correspond au nombre de visages qui, parmi les 40 exemples reconnus par l'ADL2DoL mais pas par l'ADL2DoC, sont également bien classés par l'ADB1. La table 5.1 montre que l'ADB1 reconnaît $\frac{1858}{1871} = 99{,}3\%$ des visages qui sont correctement classés à la fois par l'ADL2DoL et l'ADL2DoC. De plus, l'ADB permet de reconnaître la majeure partie des visages qu'une seule de ces deux techniques parvient à reconnaître (en moyenne $\frac{24+21}{40+30} = 64{,}3\%$). Enfin, l'ADB permet de reconnaître une part importante (28,8%) des visages qui sont conjointement mal classés par les deux versions de l'ADL2Do. Ces résultats tendent à prouver l'efficacité de l'ADB en tant que mode de combinaison de ces deux techniques, mais aussi le

	Total	**∩ADB1**
ADL2DoL ∩ ADL2DoC	1871	1858
ADL2DoL ∩ ⌉ADL2DoC	40	24
⌉ADL2DoL ∩ ADL2DoC	30	21
⌉ADL2DoL ∩ ⌉ADL2DoC	59	17
Total	2000	1920

TAB. 5.1 – *Table de contingence, contenant des valeurs sommées sur les dix partitions envisagées.*

fait que le double traitement statistique de l'ADB appliqué conjointement sur les lignes et les colonnes de l'image est plus efficace que chacun de ces deux traitements effectués séparément.

La table 5.1 montre que le taux de reconnaissance moyen de l'ADB sur les dix partitions aléatoires de la base ORL est de 96%. Si l'on compare ce résultats à ceux de la table 4.5 donnée en p. 114, on voit que l'ADB fournit sur la base ORL et pour le protocole considéré le meilleur taux moyen de reconnaissance, avec une amélioration de 0,37% par rapport à la technique d'ACP+ADL$_0$ [HLLM02].

Nous cherchons maintenant à comparer la *stabilité* de l'ADB, de l'ADL2DoL et de l'ADL2DoC. La stabilité d'un classifieur peut être caractérisée par l'impact en termes de taux de reconnaissance d'un petit changement dans les bases considérées pour l'évaluation. On construit pour cela les *boîtes à moustache* calculées à partir des dix taux de reconnaissance obtenus sur les partitions aléatoires. Ces graphes sont regroupés en figure 5.6 et donnent, pour chaque méthode, le minimum, le maximum, les trois quartiles intermédiaires (traits) et la moyenne (croix) des dix taux de reconnaissance. En gardant à l'esprit que ces graphes son construits avec 10 données seulement, on peut cependant constater que l'ADB est caractérisée par des distances interquartiles plus homogènes et réduites et semble donc être légèrement plus stable que les deux versions de l'ADL2Do dont elle est issue.

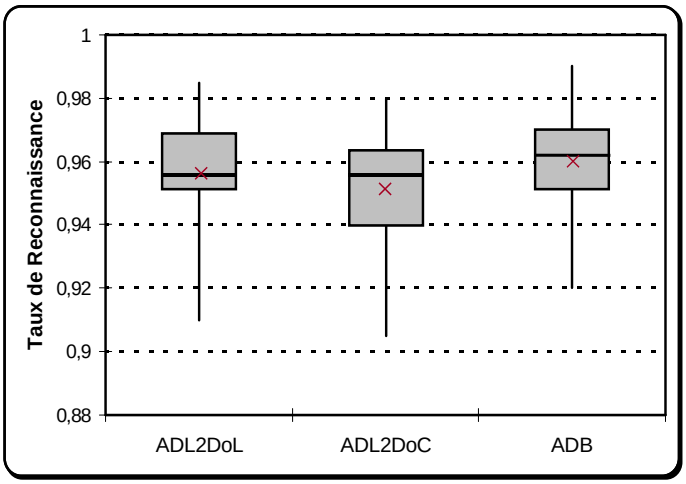


FIG. 5.6 – *Boîtes à moustaches construites depuis dix partitions aléatoires de la base ORL.*

5.2.5.2 Expérimentation menée sur FERET

L'expérience présentée dans cette section, menée sur une sous-base de FERET (*cf.* annexe A), vise à évaluer les performances de l'ADB pour la mise en correspondance d'images de personnes non enregistrées dans la base d'apprentissage. Ce contexte se retrouve dans le cadre de certaines applications d'indexation de contenu spécifique. La base d'apprentissage utilisée est la même que celle utilisée pour l'évaluation de l'ADL2Do en section 4.4.6.3. Elle contient 818 images de 152 personnes, avec au moins 4 vues par personne (*cf.* figure 5.7-a). On définit une base de connaissance et une base de test, chacune contenant 200 personnes avec une vue par personne (voir figure 5.7-b-c). Celles-ci sont tirées de FERET, mais aucune des 200 personnes n'est enregistrée dans la base d'apprentissage. L'expression faciale diffère entre la base de connaissance et la base de test. La base de test est comparée à la base de connaissance en utilisant la distance D_{L_2} au plus proche voisin. On en déduit des taux de reconnaissance qui sont donnés en figure 5.8. Celle-ci

(a) (b) (c)

FIG. 5.7 – *Extraits (a) de la base d'apprentissage, (b) de la base de connaissance et (c) de la base de test, pour la deuxième expérimentation. Toutes les images sont tirées de FERET, mais les bases de connaissance et de test ne contiennent aucun des visages enregistrés dans la base d'apprentissage.*

montre que l'ADB1 a une meilleure capacité de généralisation à des visages non enregistrés que l'ACP2D et que les deux versions issues de l'ADL2Do.

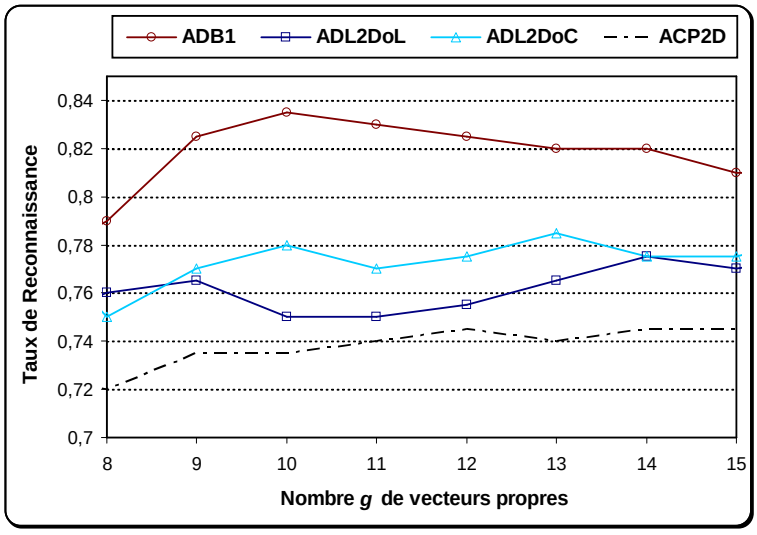


FIG. 5.8 – *Taux de reconnaissance obtenus par ADB, ADL2DoL, ADL2DoC et ACP2D, quand le nombre g de vecteurs propres retenus varie. Les taux de reconnaissance correspondent à la comparaison de la base de test avec la base de connaissance.*

5.2.5.3 Expérimentations menées sur la base AR et discussion autour des problèmes d'illumination

Les expériences présentées dans cette section sont menées sur la base AR (*cf.* annexe A) et visent à évaluer l'impact de changements dans les conditions d'illumination sur l'ADB, ainsi qu'à fournir une comparaison avec l'ACP2D et l'ADL2Do en présence de tels changements. Deux expérimentations, détaillées ci-après, sont menées. Les bases utilisées sont constituées d'images de 95 personnes dont 18 portent des lunettes, sous une pose frontale et avec peu de variations dans l'expression faciale mais avec des changements importants d'illumination, de manière à étudier de manière indépendante les effets de ces derniers (*cf.* figure 5.9). Les taux de reconnaissance obtenus sont donnés en table 5.2.

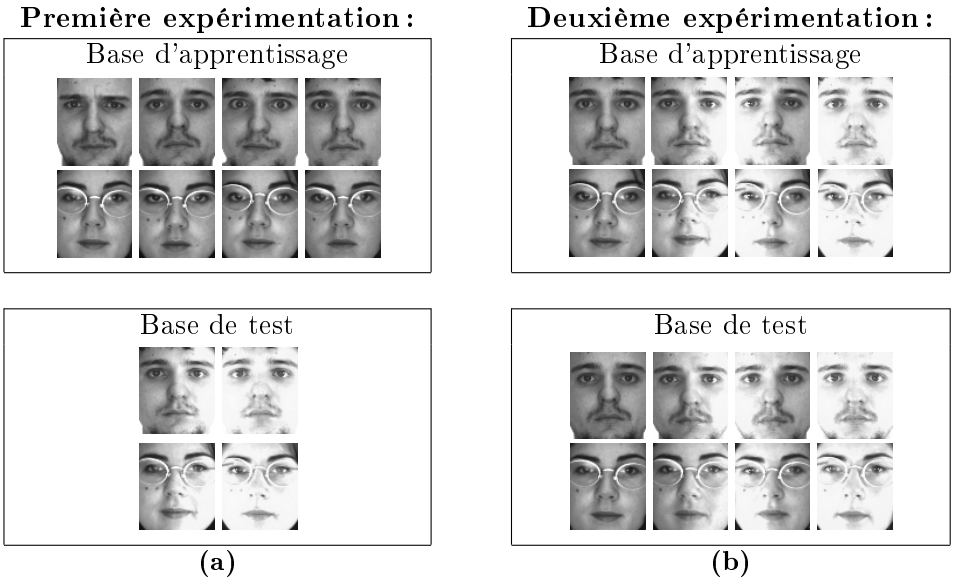


FIG. 5.9 – *Bases d'apprentissage et de test utilisées pour (a) la première expérimentation et (b) la seconde. Les images sont extraites de la base AR (cf. annexe A, p. 169).*

La première expérience consiste à utiliser pour l'apprentissage une base ne contenant que des conditions d'illumination neutres, et à lui confronter deux bases de test prises dans des conditions d'illumination très différentes (source de luminosité placée à gauche ou des deux côtés du visage). Les bases utilisées sont illustrées en figure 5.9-a. La table 5.2 montre que, dans ces conditions, l'ADB donne de bien meilleurs résultats que l'ACP2D. Les taux de reconnaissance donnés dans le tableau 5.2 sont calculés à partir du nombre de vecteurs propres donnant les meilleurs taux de reconnaissance. Il faut noter que le nombre g de vecteurs propres fournissant les meilleurs taux de reconnaissance pour l'ADL2DoC et l'ADB sont beaucoup plus importants que dans des conditions d'illumination moins inhabituelles : leurs valeurs sont de l'ordre de 22, contre 6 à 10 habituellement. Par contre, pour l'ADL2DoL et l'ACP2D, un faible nombre de composantes (de 4 à 6) suffit à garantir les meilleurs résultats de l'approche. Le critère du Lambda de Wilks de sélection (*cf.* section 4.4.4.1), uniquement basé sur l'étude de la séparation des classes de la base d'apprentissage n'a aucun moyen d'anticiper ce type de configuration et n'est pas adapté dans ce contexte. Pour déterminer la valeur optimale de g, il est possible d'utiliser une base de validation et les résultats obtenus seront d'autant meilleurs que la base de validation contiendra des types de variations proches de ceux de la base utilisée pour le test. Ainsi, nous avons testé le cas particulier où la base de validation contient des images dans les mêmes conditions que la base de test (images tirées de la seconde session de AR, menée 15 jour plus tard que la première session). Les valeurs de g obtenues sont celles permettant au système de fournir les meilleurs taux de reconnaissance. Mais cette méthodologie nécessite une analyse qualitative du contenu de la base de test, ce qui est une tâche difficile.

La deuxième expérimentation vise à comparer les performances des trois approches dans des conditions idéales, où les mêmes types de variations d'illumination sont représentés dans les bases d'apprentissage et de test. L'apprentissage est effectué à partir de quatre vues par personne collectées lors de la première session de la base AR (*cf.* annexe A). Une de ces vues est prise dans des

Expérimentation	Base de test	ADB	ADL2DoL	ADL2DoC	ACP2D
Première	gauche	78,9%	71,6%	82,1%	55,8%
	2 côtés	61,1%	45,3%	67,4%	50,7%
Deuxième	gauche	83,2%	81,1%	85,3%	77,9%
	droite	81,1%	75,8%	83,2%	76,8%
	2 côtés	90,5%	87,4%	93,7%	81,1%
	neutre	87,4%	84,2%	91,6%	86,3%

TAB. 5.2 – *Comparaison des taux de reconnaissance de l'ADB, de l'ADL2DoL, de l'ADL2DoC et de l'ACP2D pour les trois expérimentations menées sur la base AR. La base d'apprentissage diffère d'une expérimentation à l'autre. Les bases de test correspondent à différentes positions de la source lumineuse par rapport au visage (à gauche, à droite ou des deux côtés). Dans la première expérimentation, on utilise une base d'apprentissage ne contenant que des conditions d'illumination neutres, tandis que dans la deuxième expérience la base d'apprentissage contient les mêmes conditions d'illumination que les bases de test.*

conditions d'illumination neutres, les trois autres avec des changements de position de la source lumineuse (à droite, à gauche et des deux côtés). On confronte à cette base d'apprentissage trois bases de test contenant chacune une vue par personne et correspondant à ces mêmes conditions d'illumination, sauf que ces bases de test sont extraites de la seconde session AR. Il n'y a donc pas de recoupement entre les bases d'apprentissage et de test, qui sont illustrées en figure 5.9-b. On utilise également une base de test montrant une image par personne dans des conditions d'illumination neutres, afin de vérifier que la forte représentation des changements d'illumination dans la base d'apprentissage n'a pas induit de baisse des performances dans des conditions moins drastiques d'éclairage. La table 5.2 montre que le fait que la base d'apprentissage soit suffisamment représentative des conditions de prise de vue de la base de test profite à toutes les approches ; la hausse des taux de reconnaissance est particulièrement importante pour l'ACP2D et à l'ADL2DoL.

On peut remarquer que, conformément aux observations faites en section 4.4.7 sur la base de Yale, l'ADL2DoC fournit de meilleurs résultats que l'ADL2DoL, surtout en présence d'une source d'illumination placée sur le côté du visage. L'ADB fournit un taux de reconnaissance intermédiaire entre ces deux techniques. Étant donné que, dans des conditions d'illumination moins drastiques, l'ADB est plus performante (et moins coûteuse en phase de classification) que l'ADL2DoC, le choix par défaut de l'ADB reste néanmoins le plus judicieux. Par contre, si l'on ajoute en amont de la construction du modèle un module de catégorisation des conditions d'illumination, il pourrait être intéressant dans certains cas difficiles d'utiliser la technique d'ADL2Do à la place de l'ADB, afin d'optimiser les taux de reconnaissance. Cependant, la classification des conditions d'illumination est un processus difficile.

Notons qu'afin de tester l'utilité du processus d'égalisation d'histogramme mené en fin de normalisation (*cf.* annexe C), les mêmes expérimentations ont été menées sans cette étape préliminaire. La chute des taux de reconnaissance est très importante, avec une baisse des taux de reconnaissance relative de l'ordre de 50% par rapport aux images égalisées. Cela prouve l'utilité de l'opération d'égalisation d'histogramme en amont de la reconnaissance.

5.2.6 Discussion

Il est à noter premièrement que l'ADB, comme l'ADL2Do, contourne le problème de la singularité, et est généralement beaucoup moins coûteuse en termes de nombre de calculs que les *fisherfaces*, pour la construction du modèle. En effet, les tailles des matrices de dispersion sont très réduites ($w \times w$ pour la première étape de l'ADB et $h \times h$ pour la seconde contre $N \times N$ et $(N-k) \times (N-k)$ pour les deux étapes successives d'ACP et d'ADL mises en œuvre pour les *fisherfaces* (*cf.* section 3.3.3.2)).

Deuxièmement, l'ADB amène une réduction significative dans la dimensionnalité des signatures, comparé à l'ACP2D et à l'ADL2Do : la taille d'une signature issue de l'ADB est de g^2, contre hg pour l'ADL2DoL et l'ACP2D et wg pour l'ADL2DoC. Afin d'illustrer cet état de fait, prenons l'exemple des expérimentations que nous avons menées sur la base ORL, avec une résolution de 61×46 pixels. En moyenne, les meilleurs résultats de ADL2DoL étaient obtenus avec $g^* = 5$ vecteurs propres, contre $g^* = 10$ pour l'ADL2DoC, $g^* = 8$ pour l'ADB et l'ACP2D et $g^* = 39$ pour les *fisherfaces*. Les tailles des signatures associées sont donc de 488 éléments pour l'ACP2D, 460 pour l'ADL2DoC, 305 pour l'ADL2DoL, contre seulement 64 pour l'ADB et 39 pour les *fisherfaces*. Le fait que l'ADB ait engendré une réduction de taille des signatures en comparaison avec les deux versions de l'ADL2Do dont elle est issue n'a pas engendré de baisse des performances. Bien au contraire, l'ADB permet d'obtenir de meilleurs taux de reconnaissance que chacune de ces deux techniques. L'ADB permet donc de rejeter une partie du bruit des modèles, venant confirmer l'analyse que nous avions menée en section 5.2.2.2. D'un point de vue de complexité du système, cette réduction des tailles est très intéressante car elle permet d'avoir des signatures du même ordre de dimension que celles obtenues à partir de techniques 1D (généralement entre les *fisherfaces* et les *eigenfaces*).

La table 5.3 donne un classement au vu de différents critères de l'ADB et des principales techniques de projection statistique de l'état de l'art. Elle reprend le classement donné en table 4.9, en y insérant la technique d'ADB. Le rang de performance est déduit des résultats expérimentaux obtenus sur la base ORL (*cf.* section 5.2.5.1). On voit que l'ADB est très bien placée dans ce classement pour tous les critères étudiés. Notons que la technique d'ACP+ADL$_0$, qui est également très bien classée, présente le désavantage de donner des taux de reconnaissance variables en fonction des caractéristiques de la base d'apprentissage. En effet, on a notamment observé que ses performances baissent lorsque le nombre d'exemples de la base d'apprentissage augmente [CNWB05], ce qui n'est pas le cas de l'ADB.

5.2.7 Conclusion

Dans cette partie, nous avons introduit une nouvelle méthode d'extraction de signature combinant les deux versions complémentaires de l'ADL2Do. Nous avons montré qu'il s'agit d'une méthode très performante, alliant efficacement les avantages des deux techniques dont elle est issue. De plus, elle engendre une réduction dans la dimension des signatures comparé à celles de l'ADL2Do, ce qui la place en très bonne position dans le classement des techniques de l'état de

Rang	Performance	Coût de construction	Coût de classification	Coût de stockage
1	ADB	ADL2Do, ACP2D	ACP+ADL_0	ACP2D, ADL2Do
2	ADL2Do, ACP+ADL_0	ADB	*fisherfaces*, ADLD	ADB
3	ACP2D	ADLD	ADB	ACP+ADL_0
4	*fisherfaces*	*eigenfaces*	*eigenfaces*	*eigenfaces*, ADLD
5	ADLD	*fisherfaces*	ADL2Do, ACP2D	*fisherfaces*
6	*eigenfaces*	ACP+ADL_0		

TAB. 5.3 – *Classement des principales méthodes présentées dans cette section, de la plus efficace à la moins efficace, en fonction de quatre critères. Les méthodes testées sont : l'AB, l'ADL2Do (ADL2DoL et ADL2DoC) l'ACP2D [YZFY04], l'ADL dans le noyau de Huang et al. [HLLM02] (ACP+ADL_0), l'ADL Directe (ADLD) [YY01], ainsi que les techniques des eigenfaces [TP91] et des fisherfaces [BHK97].*

l'art.

5.3 Vers une approche hybride : la fusion d'experts modulaires

5.3.1 Introduction

Comme nous avons pu le constater au travers d'expérimentations présentées en section 4.4.6, la présence de variations dans l'expression faciale engendre des ambiguïtés lors de la classification des visages. En effet, il arrive qu'un classifieur mette en correspondance deux visages différents mais arborant la même expression faciale. Un changement d'expression faciale peut se répercuter de manière différente dans les différentes régions de l'image du visage : une bouche ouverte affectera plus particulièrement le bas du visage, tandis que la fermeture des yeux engendrera plus de changements dans la partie supérieure du visage. Il a été montré (*cf.* section 2.3.2) que le fait de combiner efficacement plusieurs classifieurs construits sur des régions faciales différentes peut améliorer les taux de reconnaissance, par rapport à un classifieur unique. De telles techniques, qualifiées de *modulaires* et présentées en section 2.3.2, sont conçues pour être plus robustes aux différentes sources de variation. En effet, lorsqu'un changement affecte plus particulièrement une région faciale, le classifieur correspondant devient moins performant. On espère compenser cette perte de performance par l'utilisation conjointe d'autres classifieurs, moins affectés par ce changement. C'est pourquoi nous introduisons une méthode basée sur la combinaison de plusieurs classifieurs, issus de l'ADB et entraînés indépendamment sur des régions faciales différentes : l'*Analyse Discriminante Bilinéaire Modulaire* (ADBM). Les classifieurs composant l'ADBM seront par la suite appelés *experts*.

Étant donné que cette technique repose à la fois sur la globalité du visage, et sur des régions localisées de celui-ci, elle peut être qualifiée d'*hybride*. Différents modes de combinaison sont étudiés. Ce travail a fait l'objet d'une publication dans [VGJ05b].

5.3.2 La combinaison d'experts

On peut considérer qu'il existe deux grandes familles de combinaison d'experts. Dans la première, tous les experts utilisent la même représentation des données, mais des techniques d'extraction différentes, choisies pour être complémentaires. Dans la seconde, chaque expert utilise sa propre représentation du signal d'entrée, mais l'extracteur de signature est le même. Nous nous focaliserons ici sur le deuxième cas de figure. Nous proposons une méthode de reconnaissance de visages basée sur trois experts construits par ADB, chacun d'entre eux étant entraîné sur une région faciale qui lui est spécifique.

5.3.2.1 Choix des régions faciales

Les régions faciales à partir desquelles sont construits les classifieurs sont illustrées en figure 5.10. Elles sont choisies de manière à garantir de bons résultats dans la plupart des configurations, selon les résultats de [PMS94, PG05]. De plus, le choix de ces régions faciales est en cohérence avec les études biologiques, présentées en section 1.4, qui précisent que l'œil humain tient plus compte de la partie supérieure du visage que de la partie basse pour la reconnaissance.

Le classifieur 1 est construit à partir d'une région de 75 pixels de haut et 65 pixels de large, contenant tous les éléments faciaux (région faciale globale, utilisée pour l'ADL2Do et l'ADB). L'expert 2 utilise une région de 40 pixels de haut et 65 pixels de large contenant les yeux, les sourcils et une partie du nez. L'expert 3, quant à lui, est construit à partir d'une région de 30×65 pixels, contenant uniquement les yeux et les sourcils. Nous n'avons pas retenu de région faciale

FIG. 5.10 – *Régions faciales utilisées pour construire les trois experts.*

centrée autour de la bouche, en raison des études biologiques et du fait que son apparence est très influencée par des changements d'expression faciale [PMS94]. La région la plus stable est celle des yeux (classifieur 3), mais peut être modifiée lorsque le sujet ferme les yeux par exemple. Le classifieur 3 est tolérant à des changements drastiques de la forme de la bouche, tandis que le classifieur 1 permet d'apporter l'information globale nécessaire (*cf.* section 2.3.2). Le classifieur 2, lui, est un intermédiaire entre le classifieur 1 et le classifieur 3 : moins sensible à des déformations de la bouche que l'expert 1, il véhicule plus d'information globale que le classifieur 3. Dans le contexte d'un vote à la majorité, il servira essentiellement d'arbitre, le cas échéant.

5.3.2.2 Choix de l'extracteur de signatures

Pour ses très bonnes performances et sa rapidité accrue de construction, nous choisissons de mettre en œuvre comme extracteur de caractéristiques l'ADB1 (ADB Algorithme 1). Les trois experts jouent un rôle différent et doivent permettre de représenter un niveau de détails plus ou moins important. Tandis que l'expert 1 vise à fournir une information globale (basses fréquences) du visage, les experts 2 et 3 se focalisent graduellement sur des régions de plus en plus réduites du visage, et sont conçues pour fournir de plus en plus de détails sur ces régions. Les tailles des

espaces de projection ne seront donc pas les mêmes pour les trois experts : la taille g_1 de l'expert 1 sera inférieure à g_2, elle-même inférieure à g_3.

5.3.2.3 Choix du mode de combinaison des experts

Nous avons étudié deux schémas de combinaison d'experts, que nous avons choisi de désigner par les termes d'*Agrégation d'Experts* et de *Fusion d'Experts*. Dans les deux cas, chaque classifieur $e \in \{1,2,3\}$ est préalablement construit en appliquant une ADB sur sa propre base d'apprentissage. On obtient ainsi pour chaque expert e un couple de matrices optimal (Q_e^*,P_e^*), avec un nombre de vecteurs propres g_e spécifique à l'expert concerné.

Agrégation d'experts Dans le schéma d'agrégation d'experts (*cf.* figure 5.11), la classification de chaque image-requête T se fait de la manière suivante :

- pour tout expert e, on calcule la signature $T^{(Q_e,P_e)}$ de T à l'aide de (Q_e,P_e) ;
- pour chaque expert e, on compare la signature $T^{(Q_e,P_e)}$ aux signatures $X_l^{(Q_e,P_e)}$ de sa propre base d'apprentissage : pour chaque couple (e,j) constitué de l'expert e et de la classe j, on calcule le score suivant [PG05] :

$$s^e(T,j) = \frac{\max\limits_{X_l \in \Omega_j} \left[D_{L_2}(T^{(Q_e,P_e)},X_l^{(Q_e,P_e)}) \right]^{-1}}{\sum\limits_{j=1}^{k} \max\limits_{X_l \in \Omega_j} \left[D_{L_2}(T^{(Q_e,P_e)},X_l^{(Q_e,P_e)}) \right]^{-1}} \tag{5.14}$$

où la distance D_{L_2} est donnée en équation (5.13). Ensuite, pour chaque classe $j \in \{1,\ldots,k\}$, les trois scores $s^e(T,j)$ sont agrégés pour obtenir un résultat de classification. Deux modes d'agrégation [KHDM98], à savoir le *vote à la majorité* et la *règle de la somme*, sont évalués. Le vote à la majorité consiste à assigner à l'image de test T l'identité à laquelle elle est le plus fréquemment associée. En cas d'ambiguïté (c.-à-d. si chacun des experts assigne à T une identité différente), l'expert 1 est le vainqueur.

La règle de la somme consiste à calculer, pour chaque classe j allant de 1 à k, une mesure de similarité $s(T,j)$ par sommation des scores obtenus par chaque expert :

$$s(T,j) = \sum_{e=1}^{3} s^e(T,j) \tag{5.15}$$

Nous pouvons alors définir une mesure de confiance. Notons j_1 la classe obtenant le plus haut score de similarité : $\forall j \in \{1, \ldots, k\}$, $s(T,j_1) > s(T,j)$. Dans ce contexte, l'identité j_1 est assignée à l'image T. Notons j_2 la classe obtenant le plus haut score de similarité, après j_1 :

$$\forall j \in \{1, \ldots, k\} - \{j_1\},\ s(T,j_2) > s(T,j)$$

On peut alors définir la mesure $b(T,j_1)$, qui constitue un indice de la confiance que l'on peut accorder à l'assignation de T à la classe j_1 :

$$b(T,j_1) = log\left(\frac{s(T,j_1)}{s(T,j_2)}\right) \tag{5.16}$$

Si la valeur de cette mesure est trop faible, alors il y a ambiguïté entre j_1 et j_2 pour la classification.

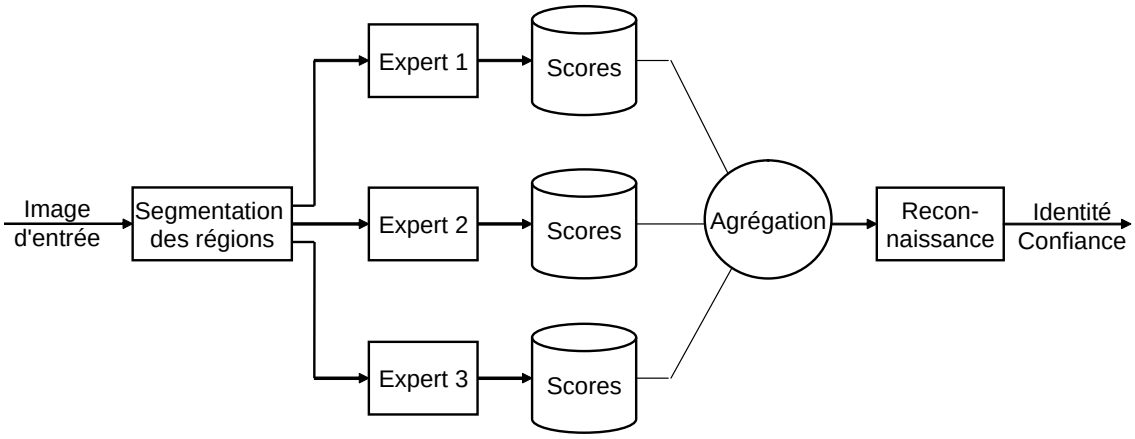


FIG. 5.11 – *Agrégation d'Experts. La classification se fait grâce à une règle permettant de combiner les sorties des trois experts.*

Fusion d'experts La fusion d'experts nécessite un apprentissage en deux temps. Après avoir construit chacun des experts indépendamment (voir section 5.3.2.3), on applique à tout $(X_l) \in \Omega$ la procédure suivante (voir figure 5.12) :

- pour tout expert e, on calcule la projection $X_l^{(Q_e,P_e)}$ de X_l sur (Q_e,P_e), selon l'équation (5.1). La matrice $X_l^{(Q_e,P_e)}$ obtenue est de taille $g_e \times g_e$;
- chacune de ces matrices $X_l^{(Q_e,P_e)}$ est transformée en un vecteur x_l^e, de longueur g_e^2, par concaténation de ses lignes ;
- les trois vecteurs $(x_l^e)_{e\in\{1,2,3\}}$ sont concaténés pour obtenir un unique vecteur x_l, de longueur $g_1^2 + g_2^2 + g_3^2$.

La longueur $g_1^2 + g_2^2 + g_3^2$ de ces vecteurs est très importante. De plus, il y a des fortes chances qu'une partie des informations provenant des différents classifieurs soit corrélée. On détermine un sous-espace de projection $\mathcal{F}$ dont les composantes sont décorrélées en appliquant une ACP sur les $(x_l)_{l=1,\ldots,N}$. Cette ACP permet de fusionner les résultats des trois experts : elle reçoit en entrée les signatures concaténées des trois classifieurs, et donne en sortie un condensé décorrélé de ces informations. On peut ici établir un parallèle avec la technique des Modèles Actifs d'Apparence (MAA) (*cf.* section 2.2.3), qui eux aussi fusionnent et décorrèlent l'information provenant de différentes sources (forme et texture) par le biais d'une ACP. La dimension des signatures est réduite, passant de $g_1^2 + g_2^2 + g_3^2$ à une taille utile m très inférieure. Enfin, tous les vecteurs x_l de la base d'apprentissage sont projetés dans $\mathcal{F}$ pour obtenir un ensemble de méta-signatures $\tilde{x_l}$, sous la forme de vecteurs de longueur m.
Lorsqu'une image-requête T est présentée au système, on applique la même procédure de projection en trois étapes que pour les images de la base d'apprentissage. On obtient ainsi sa méta-signature $\tilde{t}$. Puis $\tilde{t}$ est comparée aux méta-signatures de la base d'apprentissage $\tilde{x_l}$ en utilisant la distance Euclidienne au plus proche voisin.

5.3.3 Évaluation de la méthode proposée

Dans cette partie, nous évaluons les performances de la méthode proposée en utilisant une sous-base de la base Asian Face Image Database PF01 (*cf.* annexe A) contenant 75 personnes, sous des conditions d'illumination neutres et ne portant pas de lunettes. La méthode proposée

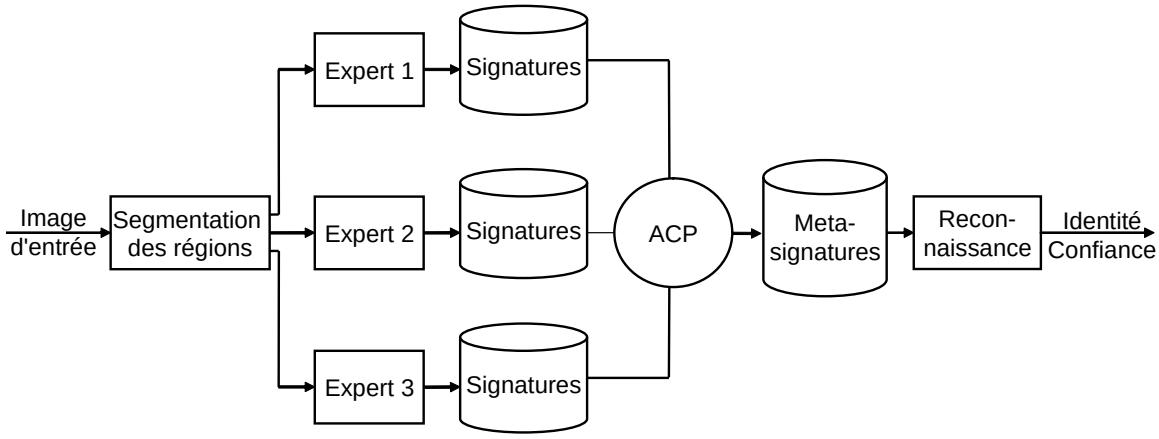


FIG. 5.12 – *Fusion d'Experts. Les signatures fournies par les trois experts sont combinées en utilisant une ACP, pour obtenir une unique méta-signature.*

est comparée à chacun des trois experts construits indépendamment, et à la méthode dite des *Modular Eigenspaces* [PMS94] (voir section 2.3.2).

L'objectif de la première expérimentation est de tester si, comme nous l'espérons, l'ADBM est plus performante que les autres méthodes, en présence de changements drastiques dans l'expression faciale. La base d'apprentissage (*cf.* figure 5.13-a) contient quatre vues par personne, sous des poses presque frontales et des expressions faciales neutres. Trois bases de test sont considérées (*cf.* figure 5.13-b) : la première contient une vue par personne exprimant la colère, la seconde contient des visages souriants tandis que la troisième montre des visages surpris. Entre les bases d'apprentissage et de test, il y a des variations drastiques dans l'expression faciale. Nous

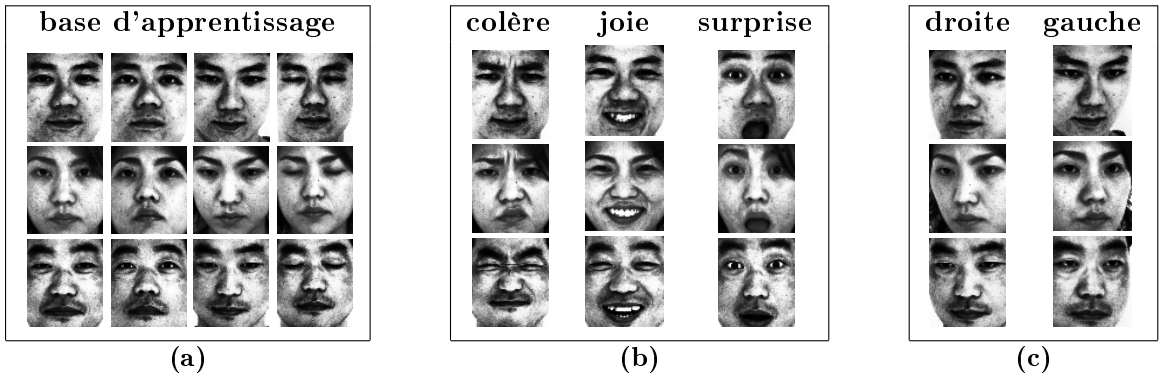


FIG. 5.13 – *Extraits (a) de la base d'apprentissage, (b) des bases de test utilisées pour la première expérimentation et (c) des bases de test utilisées pour la seconde expérimentation.*

avons également conçu une seconde expérimentation pour vérifier que l'ADBM n'est pas moins performante que les autres méthodes en présence d'autres sources de dissimilarités, comme des changements dans la pose de la tête par exemple. La base d'apprentissage est la même que pour la première expérimentation ; les deux bases de test, illustrées en figure 5.13(c), diffèrent dans la pose de la tête. Chacune des cinq bases de test est comparée à la base d'apprentissage en utilisant une distance Euclidienne au plus proche voisin.

5.3.3.1 Effet du paramètre g

Pour la base d'apprentissage considérée, les g_e optimaux (fournissant les meilleurs taux de reconnaissance) sont respectivement 14, 15 et 17 pour les experts 1, 2 et 3. Cette constatation vient confirmer l'analyse menée en section 5.3.2.2, selon laquelle de l'expert 1 à l'expert 3 le niveau de détails nécessaire croît.

5.3.3.2 Comparaison de ADBM et de l'ADB

La figure 5.14 donne les taux de reconnaissance obtenus par chacun des trois experts, et par l'algorithme de fusion des experts (*cf.* figure 5.12). On constate que, en présence de variations dans la pose de la tête, ou bien quand l'expression faciale engendre de fortes modifications dans l'aspect des yeux et des sourcils (base « colère »), l'expert 1, construit à partir du visage entier, est plus performant que chacun des deux autres experts pris séparément. Néanmoins, quand l'expression faciale engendre des changements d'aspect drastiques de la région basse du visage (bases « sourire » et « surprise »), l'expert 3 est le plus performant. Dans tous les cas, l'ADBM avec fusion d'experts donne de meilleurs résultats de classification que chacun des experts, pris séparément, avec une amélioration moyenne des taux de reconnaissance de 3,9% sur les cinq bases de test, par rapport au taux de reconnaissance moyen des trois experts (soit en moyenne une amélioration de 15 visages sur 375). Cette amélioration peut être considérée comme faible ; cependant l'ADBM est plus *stable* que chacun des experts pris séparément, puisqu'il est systématiquement plus performant que les trois, alors que ceux-ci peuvent se dépasser les uns les autres en fonction des bases.

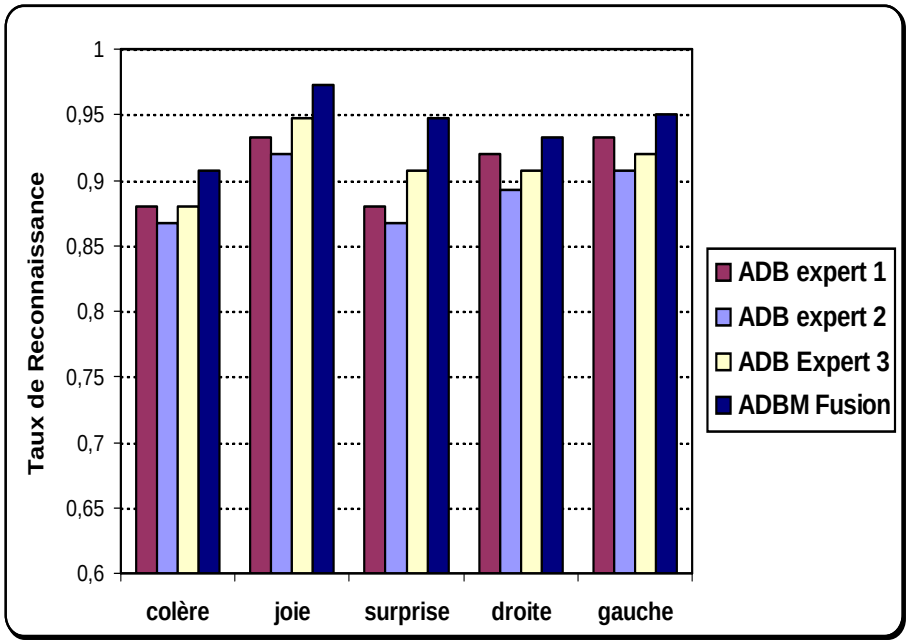


FIG. 5.14 – *Taux de reconnaissance comparés des trois experts utilisés séparément, et de la fusion des trois experts.*

5.3.3.3 Comparaison de l'ADBM et des Modular Eigenspaces

La figure 5.15 permet de comparer les taux de reconnaissance de l'ADB et des Modular Eigenspaces avec une combinaison par agrégation d'experts : Vote à la Majorité (VM) ou Règle de la Somme (RS), et l'ADB avec Fusion d'Experts (FE). Cette figure nous apprend que l'ADBM est plus performante que la méthode des Modular Eigenspaces (avec respectivement pour le vote à la majorité et la règle de la somme des améliorations de 6,14% (23 visages) et 6,42% (24 visages) des taux de reconnaissance). Notons également que l'ADBM avec Fusion d'Experts amène une amélioration de plus de 9% (soit 34 visages sur 375) par rapport à chacune des deux versions des Modular Eigenspaces.

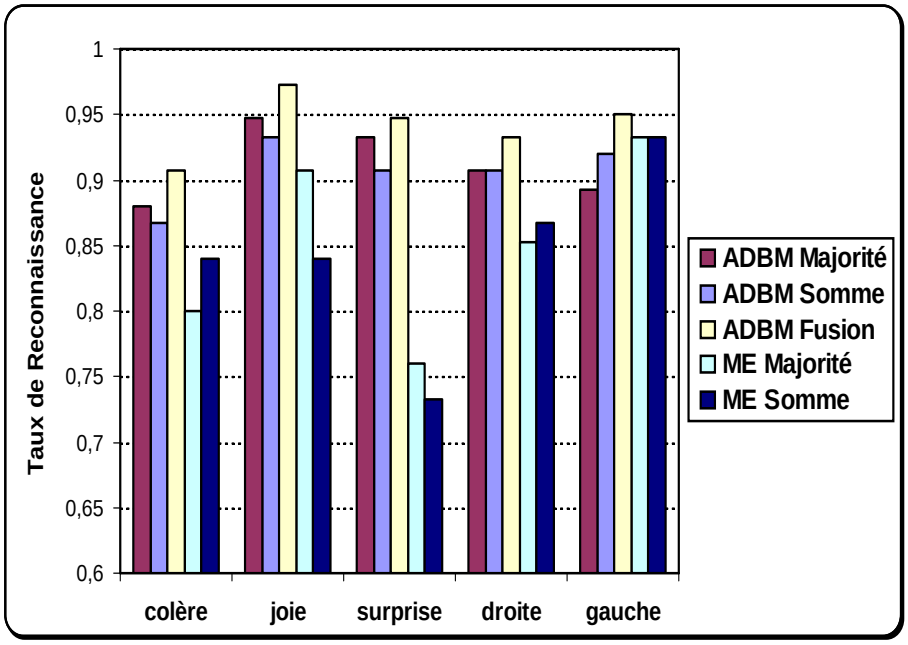


FIG. 5.15 – *Taux de reconnaissance de l'ADBM et des Modular Eigenspaces (ME) utilisant : le Vote à la Majorité (Majorité) et la Règle de la Somme (Somme), et de l'ADBM avec Fusion d'Experts (ADBM Fusion).*

5.3.3.4 Sélection du mode de combinaison des experts

Les résultats expérimentaux donnés en figure 5.15 montrent que, pour l'ADB comme pour les *eigenfaces*, le Vote à la Majorité semble donner en moyenne des résultats légèrement meilleurs que ceux obtenus à l'aide d'une Règle de la Somme, ce qui est logique dans la mesure où la somme est moins robuste que le vote à la majorité (le vote à la majorité connaît son *point d'effondrement* (*cf.* définition en note de bas de page, p. 75) à 50%, contre 0% pour la somme). Le schéma de fusion des experts est plus performant : pour l'ADBM, on enregistre une amélioration des taux de reconnaissance de 3% et de 3,52%, respectivement, sur les algorithmes de vote à la majorité et de règle de la somme.

5.3.3.5 Évaluation de la mesure de confiance

La figure 5.16 donne les taux de reconnaissance de l'ADBM et des *Modular Eigenspaces*, construites sur la base « surprise » avec la règle de la somme, en fonction du ratio de visages rejetés à cause d'une mesure de confiance (donnée en équation (5.16)) trop basse. Par exemple, si l'on rejette 10,67% des visages-requêtes, soit huit visages au total, l'ADBM permet d'atteindre un taux de reconnaissance de 95,5%, contre moins de 79,5% pour les *modular eigenspaces*. Ce graphe montre l'adéquation de la mesure de confiance proposée avec le but recherché. En effet, ce critère nous permet de rejeter majoritairement des visages qui seraient, s'ils étaient conservés, mal classés. Par exemple, pour l'ADBM, rejeter 10,7% des visages-requêtes permet de rejeter plus de 57% des visages mal classés. Néanmoins, cette mesure ne nous semble pas assez précise pour être utilisée dans un contexte en monde ouvert, où son rôle consisterait à filtrer de manière fiable les visages selon leur appartenance ou non à la base d'apprentissage.

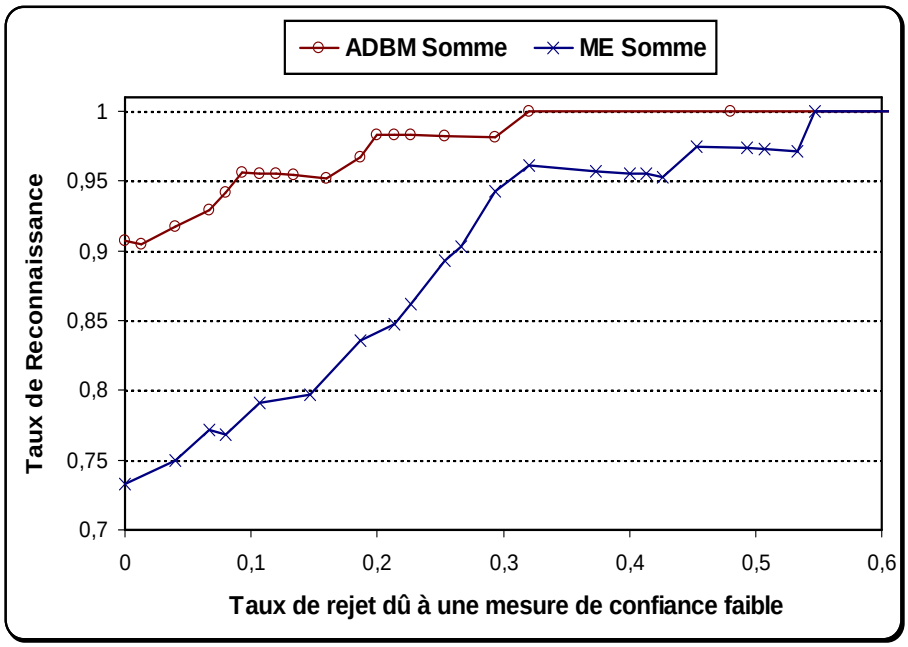


FIG. 5.16 – *Taux de reconnaissance de l'ADBM et des Modular Eigenspaces, calculés sur la base de test « surprise », en fonction du nombre d'images rejetés à cause d'une mesure de confiance trop faible.*

5.3.4 Discussion

Dans cette partie, nous avons présenté une nouvelle technique de reconnaissance de visages qui, de par sa modularité, offre une robustesse accrue aux changements d'expressions faciales tout en garantissant de très bonnes performances en présence d'autres sources de variations telles que les changements de pose. Néanmoins, pour beaucoup d'applications (telles que la vidéosurveillance), on ne travaille pas en monde fermé. En d'autres termes, un visage-requête peut ne pas être préalablement enregistré dans la base de connaissance. Il faut alors être capable de décider si un visage-requête est ou non enregistré dans cette base de connaissance. Cette phase préliminaire de filtrage des visages est souvent critique : ainsi, dans certaines applications de

vidéosurveillance, le fait qu'un visage soit reconnu comme appartenant à la base d'apprentissage peut parfois suffire à déclencher une alerte. La mesure de confiance présentée dans le cadre de l'ADBM est cohérente au sens où, dans l'espace défini par l'ADB, elle permet de rejeter en majorité des visages qui sont les plus éloignés de leur classe d'appartenance. Néanmoins, cette mesure ne nous paraît pas assez robuste pour servir de base à un outil de décision. Par la suite, nous proposons donc d'autres voies.

5.4 Classification des signatures par Réseaux de Fonctions à Base Radiale Normalisés

5.4.1 Introduction

Nous avons introduit plus tôt dans ce chapitre une nouvelle méthode d'extraction de signature, appelée Analyse Discriminante Bilinéaire (ADB) qui, en combinant efficacement les deux versions de l'ADL2Do, prend en compte l'information bidimensionnelle provenant des images. Nous avons évalué les performances de cette technique, et nous l'avons comparée aux autres techniques usuelles basées sur la projection statistique. Les résultats expérimentaux ont prouvé sur différentes bases internationales que l'ADB est plus performante que les techniques usuelles d'ADL basées sur une représentation 1D des visages, et que l'ACP2D. Nous avons choisi pour cette évaluation d'utiliser une mesure de dissimilarité, et montré pour cela l'efficacité de la distance Euclidienne avec une règle d'affectation au plus proche voisin. Néanmoins, la règle du plus proche voisin est très coûteuse en termes de temps de calcul lors de la classification, et peut être influencée par la présence d'*observations aberrantes* (voir section 3.4.3).

Dans cette section, nous proposons de remplacer la mesure de dissimilarité par un *Réseau de Fonctions à Base Radiales Normalisé* (RFBRN) [MD89] pour la classification des signatures issues de l'ADB. Un RFBRN constitue en effet un outil de classification moins coûteux en termes de temps de calcul, non linéaire et donc permettant de définir des hypersurfaces de séparation plus complexes. De plus, ce type de réseaux fournit une estimation des probabilités d'appartenance à chaque classe, ce qui nous permet de mieux approcher la distribution des classes dans l'espace défini par l'ADB, et de dériver facilement des règles de décision concernant l'appartenance ou non d'un visage à la base d'apprentissage, c'est-à-dire de travailler en monde ouvert. L'influence d'éventuelles observations aberrantes est de plus réduite. Les RFBRN se caractérisent par une très bonne capacité de généralisation et de bonnes performances en présence de données de grandes dimensions [Nel01, PS04], ce qui les rend particulièrement adaptés à la tâche de reconnaissance de visages. Si les RFBR standard ont déjà été utilisés pour la reconnaissance de visages [TFV98, FTV99, EWLT02, WJHT04], et ce avec succès (*cf.* table 2.1), c'est à notre connaissance la première fois que le RFBRN sont utilisés dans ce contexte. Ce travail a fait l'objet d'une publication dans [VGJ05c].

5.4.2 Les Réseaux de Fonctions à Base Radiale

Les Réseaux de Fonctions à Base Radiale (RFBR) se sont imposés comme variante des Réseaux de Neurones Artificiels à la fin des années 1980. Cependant, les fondements de cette technique sont enracinés dans des méthodes de reconnaissance de formes bien plus anciennes, comme par exemple les fonctions de potentiel, l'approximation de fonctions, le *clustering* (*cf.* note de bas de page n° 8 en p. 40), l'interpolation par splines ou les modèles de mélange [TG74].

5.4.2.1 Topologie du réseau

Un Réseau de Fonctions à Base Radiale (RFBR) est un réseau de neurones à deux couches (voir figure 5.17). Les neurones de sortie effectuent une combinaison linéaire des fonctions non linéaires fournies par les cellules de la couche cachée. La valeur de sortie est différente de zéro seulement si le signal d'entrée se situe dans une région bien localisée de l'espace des variables.

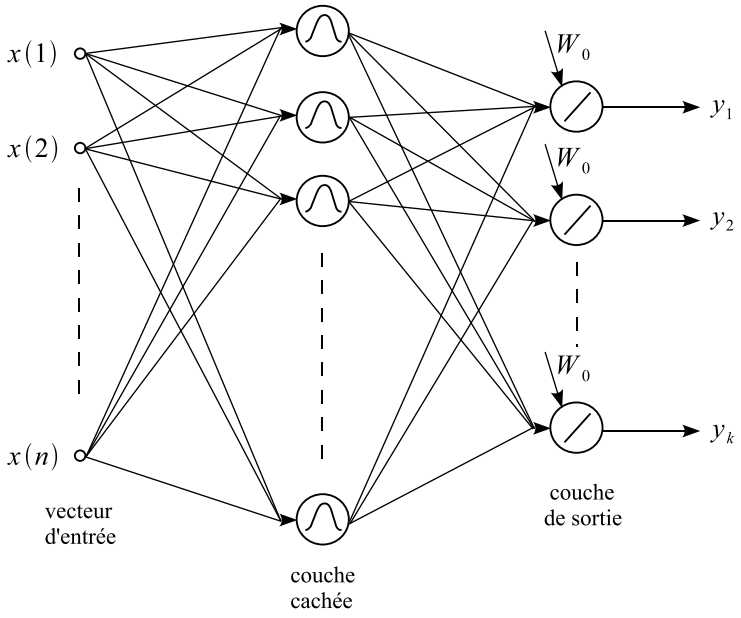


FIG. 5.17 – *Architecture d'un Réseau de Fonctions à Base Radiale (RFBR).*

Chaque neurone i de la couche cachée applique une fonction non linéaire, appelée *Fonction à Base Radiale* (FBR) et notée R_i, sur le signal d'entrée. Le terme « Fonction à Base Radiale » désigne des fonctions symétriques radialement, c'est-à-dire qu'à chacune de ces fonctions est associé un centre et que la valeur de la fonction est la même pour toutes les entrées situées à une même distance de ce centre. Les FBR sont de la forme :

$$\begin{array}{rcccc} R_i & : & \mathbb{R}^n & \rightarrow & \mathbb{R} \\ & & x & \mapsto & K(\|x - C_i\|) \end{array}$$

où $x \in \mathbb{R}^n$ est le signal d'entrée, et C_i est le centre de la $i^{\text{ème}}$ fonction à base radiale. Bien que plusieurs familles de telles fonctions existent, la plus couramment utilisée est de type Gaussien :

$$R_i(x) \quad = \quad e^{(x-C_i)^T \Sigma_i^{-1} (x-C_i)} \tag{5.17}$$

où Σ_i désigne la matrice de covariance de la $i^{\text{ème}}$ FBR . Plus le vecteur d'entrée x est proche du centre C_i de la $i^{\text{ème}}$ FBR , plus la sortie $R_i(x)$ du $i^{\text{ème}}$ neurone caché est élevée.

Les neurones de sortie fournissent une combinaison linéaire des sorties des r neurones cachés. Si l'on note W_0 le biais du système, la réponse du $j^{\text{ème}}$ neurone de sortie est :

$$y_j(x) \quad = \quad W_0 + \sum_{i=1}^{r} W_{j\,i} R_i(x) \tag{5.18}$$

Supposons que l'on dispose d'une base d'apprentissage constituée de N observations x_l. Le système peut être résumé comme suit :

$$Y = WR \tag{5.19}$$

où les éléments de la matrice $W \in \mathbb{R}^{k\times(r+1)}$ sont les $W_{j\,i}$ et, $\forall j \in \{1 \dots k\}$,$W_{j\,0} = W_0$. La matrice $R \in \mathbb{R}^{(r+1)\times N}$ contient les éléments $R_i(x_l)$, avec $R_0(x_l) = 1$, ceci pour tout l allant de 1 à N. Les éléments de $Y \in \mathbb{R}^{k\times N}$ sont les $y_j(x_l)$.

5.4.2.2 Propriétés

Les RFBR sont caractérisés par les localisations (centres) et par les hypersurfaces d'activation de leurs neurones cachés. Dans le cas du modèle Gaussien présenté ci-dessus ces caractéristiques sont modélisées par les deux paramètres C_i et Σ_i. Dans le cas Gaussien général, l'hypersurface est une hyperellipsoïde et se réduit à une hypersphère si la matrice de covariance Σ_i est diagonale, avec égalité des éléments diagonaux. Dans le cas général d'une hyperellipsoïde, l'influence de chaque neurone caché (FBR) décroît selon la distance de Mahalanobis à son centre. En d'autres termes, les exemples situés à une distance de Mahalanobis trop importante du centre n'activent pas le neurone caché correspondant tandis que, lorsque l'exemple coïncide avec le centre, l'activation est maximale. Si les centres des Gaussiennes sont suffisamment éloignés (au sens des distances de Mahalanobis, c.-à-d. de leurs dispersions), on peut considérer que les FBR Gaussiennes sont quasi-orthogonales : leur produit est proche de zéro.

5.4.2.3 Initialisation

Le nombre, la position et les domaines d'influence des FBR ont un impact important sur les performances du réseau [BG94]. Dans cette partie, nous nous intéressons aux RFBR dits « statiques », dont le nombre de FBR est fixé tout au long du processus d'apprentissage. On peut considérer qu'il existe trois grandes familles de stratégies d'initialisation des paramètres du réseau :

1. les centres sont sélectionnés au hasard parmi les observations de la base d'apprentissage [BL88], puis les domaines d'influence sont ajustés en utilisant les distances entre *clusters* (*cf.* note de bas de page n° 2.2.2.7 en p. 36) ;
2. les paramètres sont initialisés à l'aide d'un algorithme de *clustering* non supervisé [MD89, TFV98] ;
3. les paramètres sont initialisés selon une procédure supervisée [PG90b, EWLT02].

Les méthodes de type 1. sont peu performantes dans des espaces de grandes dimensions, et les techniques non supervisées peuvent, dans certains cas, converger vers un optimum local peu performant [MH98]. Les techniques d'initialisation supervisées, elles, ont fait leurs preuves dans le domaine de la classification de données de grandes dimensions, et notamment pour la classification de visages selon leurs signatures obtenues par la technique des *fisherfaces* [EWLT02] (*cf.* section 2.2.4).

5.4.2.4 Apprentissage des paramètres du réseau

Les RFBR sont le plus souvent entraînés de manière supervisée. La matrice des sorties désirées (matrice-cible), notée S et de taille $k \times N$, est connue. C'est à partir de cette matrice que l'on va chercher à optimiser les paramètres $\theta = (\theta_{i\,j})_{\substack{i=1,\dots r\\ j=1,\dots,k}}$, où $\theta_{i\,j} = \{C_i, \Sigma_i, W_{j\,i}\}$ du réseau.

L'apprentissage des trois paramètres peut être soit simultané, soit en deux étapes (les positions et zones d'influence des neurones cachés étant ajustés dans une première phase, et la matrice des poids dans une seconde étape).

Algorithme de remise à jour simultanée On cherche à minimiser la fonction de coût suivante :

$$E = \sum_{l=1}^{N} E^l = \frac{1}{2}\sum_{l=1}^{N}(S_{.l} - Y_{.l})^T(S_{.l} - Y_{.l}) \tag{5.20}$$

où $Y_{.l}$ (respectivement $S_{.l}$) est la $l^{\text{ème}}$ colonne de la matrice des sorties obtenues Y (respectivement de la matrice des sorties désirées S). À chaque époque e et pour chaque exemple x_l, les paramètres sont ajustés selon :

$$C_i = C_i - \xi_c \Delta C_i^l \tag{5.21}$$

$$\Sigma_i = \Sigma_i - \xi_\Sigma \Delta \Sigma_i^l \tag{5.22}$$

$$W_{j\,i} = W j\,i - \xi_W \Delta W_{j\,i}^l \tag{5.23}$$

où ΔC_i^l, $\Delta \Sigma_i^l$ et $\Delta W_{j\,i}^l$, dont les *taux d'apprentissage* associés sont respectivement ξ_c, ξ_Σ et ξ_W, sont calculés par le biais d'un algorithme de descente du gradient, non-linéaire. Cet algorithme peut théoriquement fournir une estimation optimale des paramètres du modèle. Mais il comporte un certain nombre de désavantages. Tout d'abord, il est très coûteux. De plus, il nécessite l'ajustement de trois taux d'apprentissage ξ_c, ξ_Σ et ξ_W, ce qui est une tâche difficile.

Algorithme en deux étapes L'apprentissage en deux étapes offre une alternative très intéressante à l'algorithme de remise à jour simultanée. Dans une première phase, les paramètres du réseau sont adaptés par le biais de l'une des stratégies, supervisées ou non, utilisées pour leur initialisation. Puis, les poids sont ajustés de manière supervisée par le biais d'un algorithme des moindres carrés. Pour des centres C_i et des dispersions Σ_i fixés, la minimisation de l'erreur de coût quadratique donnée en équation (5.20) est obtenue pour :

$$W = SR^+ \tag{5.24}$$

où $R^+ = (R^T R)^{-1} R^T$ est la pseudo-inverse de Moore-Penrose qui, pour des raisons de stabilité numérique, est généralement estimée par le biais de Décompositions en Valeurs Singulières (*cf.* définition 3.1 p. 69). Il faut noter que, parmi les algorithmes en deux étapes, on compte les algorithmes dits *hybrides*, au sens où les paramètres de la couche cachée sont ajustés par le biais non-linéaire d'un algorithme de gradient, tandis que la matrice des poids est réestimée selon la technique linéaire des moindres carrés.

5.4.2.5 Classification

Les RFBR permettent de modéliser les catégories de données avec un faible nombre de paramètres, ce qui rend la phase de classification beaucoup moins coûteuse qu'avec une règle au plus proche voisin. Généralement, pour une tâche de classification, le réseau est construit de manière à ce que chaque neurone de sortie corresponde à une classe. Lorsqu'un exemple t doit être classé, on calcule la sortie associée de la couche cachée $R(t) = [1, R_1(t), R_2(t), \ldots, R_r(t)]^T$, puis son vecteur de sortie $y(t) = [y_1(t), y_2(t), \ldots, y_n(t)]^T$, tel que $y(t) = WR(t)$. On assigne

enfin à t la classe correspondant à l'indice j de l'élément $y_j(t)$ dont la valeur est optimale dans $y(t)$. Généralement, l'apprentissage du réseau est mené de telle manière que la valeur optimale à rechercher est le maximum.

5.4.2.6 Comparaison avec les réseaux Perceptron Multi-Couches

La décomposition de l'algorithme d'apprentissage en deux phases constitue l'un des avantages principaux des RFBR sur les réseaux Perceptron Multi-Couches (PMC), car elle permet généralement aux premiers de jouir d'un apprentissage beaucoup plus rapide. De plus, la décomposition des tâches globales en une combinaison de sous-tâches locales (par le biais de l'utilisation de FBR localisées), permet une interprétation simple du réseau, à l'opposé du mode de fonctionnement de type « boîte noire » du PMC. On peut noter de plus que les RFBR ont d'excellentes capacités d'approximation non-linéaire [PS91, PG90a], ce qui leur permet de modéliser des applications complexes, que des réseaux PMC ne peuvent approximer qu'au prix d'un nombre de couches cachées très important [Hay94].

A contrario, les désavantages des RFBR proviennent essentiellement de l'utilisation de distances dans la définition des Fonctions à Base Radiale, ce qui peut poser problème si les variables observées correspondent à des échelles de mesure différentes, sont très corrélées ou peu informatives.

5.4.2.7 Les Réseaux de Fonctions à Base Radiale Normalisés

Les Réseaux de Fonctions à Base Radiales Normalisés (RFBRN) ont été introduits en 1989 par Moody et Darken [MD89]. La spécificité des RFBR Normalisés est que la sortie de chaque neurone FBR est normalisée par l'activité totale de la couche cachée. La *consistance* universelle et les taux de convergence des RFBRN ont été étudiés par Xu *et al.* [XKY94]. La normalisation utilisée rapproche les RFBRN des classifieurs Bayésiens (*cf.* section E.2, p. 192 de l'annexe E), puisqu'ils permettent d'estimer les probabilités d'appartenance à chacune des classes [GN00]. Aussi les RFBRN sont-ils très efficaces pour la classification [JS93, RMRG97, Bug98], et présentent une bonne capacité de généralisation y compris lorsque les données sont de grandes dimensions [Nel01]. Toutes ces caractéristiques rendent les RFBRN particulièrement adaptés à la reconnaissance de visages. De plus, on peut facilement dériver des estimations des probabilités conditionnelles en sortie des règles de décision concernant l'appartenance des visages-requêtes à la base de connaissance, et ainsi étendre notre approche à des applications en monde ouvert.

Dans les RFBRN, les fonctions d'activation $R_i(x)$ sont redéfinies comme suit :

$$R_i(x) \quad = \quad \frac{e^{(x-C_i)^T \Sigma_i^{-1} (x-C_i)}}{\sum_{j=1}^{r} e^{(x-C_j)^T \Sigma_j^{-1} (x-C_j)}} \tag{5.25}$$

On peut remarquer que $R_i(x)$ est une mesure de la contribution du $i^{\text{ème}}$ neurone caché à la sortie associée à l'observation x. Étant donné que, de plus, les $R_i(x)$ sont positifs et tels que $\sum_{i=1}^{r} R_i(x) = 1$, nous pouvons les interpréter comme étant liés aux probabilités $\mathbb{P}[i/x]$ que l'exemple x appartienne au domaine d'influence de la $i^{\text{ème}}$ FBR. Vu que la $j^{\text{ème}}$ sortie du réseau est une estimation de la probabilité *a posteriori* que l'observation x appartienne à la classe Ω_j :

$$y_j(x) = W_0 + \sum_{i=1}^{r} W_{j\,i} R_i(x) \simeq \mathbb{P}[\Omega_j/x] \tag{5.26}$$

les poids $W_{j\,i}$ peuvent être interprétés comme des estimations des probabilités *a posteriori* $\mathbb{P}[\Omega_j/i]$ d'association de la $i^{\text{ème}}$ FBR et de la classe Ω_j.

5.4.3 La méthode proposée

Nous avons vu en section 2.2.4 que les RFBR standard ont déjà été utilisés avec succès par Er *et al.* [EWLT02] pour la classification de signatures de visages obtenues par la technique des *fisherfaces*. Leurs résultats expérimentaux ont mis en lumière les très bonnes performances de leur approche (*cf.* table 2.1). L'utilisation d'un RFBR est donc efficace pour classer des données projetées linéairement dans un espace conçu pour être discriminant. Cependant, leur algorithme est très coûteux, et aucune solution en monde ouvert n'est proposée (voir section 2.2.4).

Nous suggérons de mettre en œuvre un RFBR Normalisé en aval de l'ADB. Nous pourrions utiliser l'ADB Modulaire au lieu de l'ADB comme extracteur de signature, mais des résultats préliminaires ont montré que cela revient à complexifier le problème (trois signatures par personne au lieu d'une) et à ralentir la classification, pour un gain insignifiant en termes de taux de reconnaissance. Ainsi, les signatures issues de l'ADB sont classées de manière non-linéaire, ce qui nous permet de définir des hypersurfaces de séparation plus complexes, si besoin. Le fait de normaliser les FBR permet la définition de règles simples pour travailler efficacement en monde ouvert.

5.4.3.1 Mode de mise en œuvre

Les matrices-signatures de la base d'apprentissage $X_l^{(Q,P)}$ sont transformées en vecteurs x_l par concaténation de leurs lignes, et ce sont ces vecteurs x_l qui constituent les signaux d'entrée du réseau de FBR. Nous choisissons d'utiliser comme FBR des Gaussiennes dont les domaines d'activation sont *hypersphériques*. Les $R_i(x_l)$ sont donc définis comme suit :

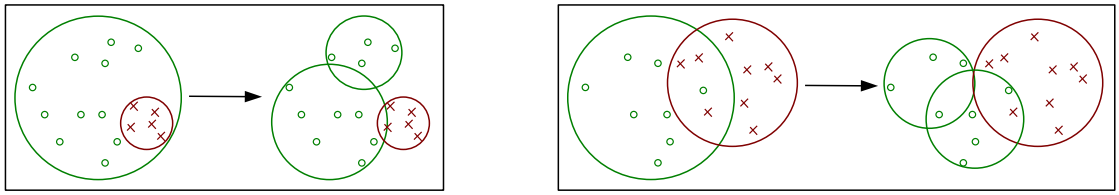
$$R_i(x) = \frac{e^{-\frac{1}{2\sigma_i^2}(x_l - C_i)^T(x_l - C_i)}}{\sum_{j=0}^{r} e^{-\frac{1}{2\sigma_j^2}(x_l - C_j)^T(x_l - C_j)}} \tag{5.27}$$

où les σ_i sont les valeurs des éléments diagonaux des $\Sigma_i = \sigma_i I$, où I est la matrice identité. Nous pourrions modéliser les *clusters* par le biais d'hyperellipsoïdes, néanmoins cela impliquerait l'estimation d'un nombre beaucoup plus important de paramètres, pour un gain généralement limité en termes de taux de reconnaissance, voire une baisse de ces derniers. Le modèle hypersphérique est très utilisé pour la classification de données de grandes dimensions [BGS05].

5.4.3.2 Initialisation

Nous avons vu en section précédente que les stratégies d'initialisation au hasard et non supervisée sont d'un intérêt limité dans le contexte de la reconnaissance de visages. C'est pourquoi nous avons choisi d'utiliser des techniques d'initialisation supervisées. Nous procéderons à l'évaluation de deux d'entre elles.

Dans un premier temps, nous avons testé la stratégie itérative introduite par Er *et al.* dans [EWLT02] et illustrée en figure 5.18. À l'initialisation, chaque classe se voit assigner un FBR, dont l'hypersphère est ajustée de manière à contenir tous les exemples provenant de cette classe. Par conséquent, on peut considérer que chaque FBR définit un *cluster*. Puis, les deux critères suivants sont passés en revue : 1) *le critère d'inclusion* : si le *cluster j* est entièrement inclus dans le *cluster k*, alors le *cluster k* doit être divisé en deux clusters ; 2) *le critère d'ambiguïté* : si le *cluster j* contient beaucoup de données provenant du *cluster k*, alors le *cluster j* doit être divisé en deux *clusters*. Cette procédure est répétée jusqu'à ce qu'aucun exemple issu de la base d'apprentissage ne satisfasse l'un de ces deux critères.



(a) **(b)**

FIG. 5.18 – *Initialisation des hypersphères des RFBRN. Illustrations en deux dimensions (a) du critère d'inclusion et (b) du critère d'ambiguïté.*

Nous avons dans un second temps évalué une technique d'initialisation plus simple, qui consiste à assigner à chaque classe un nombre fixé de FBR. Si le nombre de FBR par classe est fixé à un, alors le centre C_i du $i^{\text{ème}}$ FBR, associé à la $j^{\text{ème}}$ classe Ω_j, est le centroïde $\overline{x_j}$ de Ω_j, et son écart-type σ_i définit le rayon minimum de l'hypersphère contenant tous les exemples issus de la classe Ω_j :

$$\sigma_i = \sqrt{\max_{x_l \in \Omega_j} \|x_l - C_i\|_2} \tag{5.28}$$

Si l'utilisation de plus d'un FBR par classe est nécessaire, les FBRs additionnels sont ajoutés au hasard à l'intérieur de l'hypersphère définie par le premier FBR.

Que l'on utilise l'une ou l'autre de ces deux stratégies d'initialisation, nous appliquons une étape postérieure de réajustement des FBR, de manière à fournir un bon compromis entre *spécialisation* et *généralisation*. En effet, les hypersphères doivent être suffisamment séparées pour éviter toute ambiguïté lors de la reconnaissance de vues très proches de la base d'apprentissage (spécialisation). Pour autant, il faut éviter une « sur-spécialisation » à la base d'apprentissage (*surapprentissage*) qui pourrait engendrer des baisses des taux de reconnaissance du système dès qu'un petit changement survient entre la base d'apprentissage et la base de test. Afin d'éviter ce phénomène, on peut être amené à ne pas strictement maximiser la distance entre classe, voire à tolérer un certain chevauchement de celles-ci. Nous choisissons pour satisfaire ce compromis d'utiliser un réajustement basé conjointement sur une mesure de la dispersion à l'intérieur des classes et les distances entre classes différentes. L'ajustement proposé, inspiré de [EWLT02], est détaillé ci-après. Notons

$$d_i^W = \max_{x_l \in \omega_i} \|x_l - C_i\|_2 \tag{5.29}$$

où l'ensemble ω_i est constitué de l'ensemble des exemples inclus dans l'hypersphère délimitée par le $i^{\text{ème}}$ FBR, et :

$$d_i^B = \min_{C_p \in \omega_i'} \|C_p - C_i\|_2 \tag{5.30}$$

où ω_i' est l'ensemble des centres des FBRs, à l'exclusion de C_i. L'ajustement proposé, inspiré de [EWLT02], est :

$$\sigma_i^W = \frac{d_i^W}{\sqrt{|\log(\beta)|}}, \qquad \sigma_i^B = \mu d_i^B \tag{5.31}$$

$$\sigma_i = \max(\sigma_i^W, \sigma_i^B) \tag{5.32}$$

où le paramètre μ peut être estimé comme suit :

$$\mu \approx \frac{\sum_{j=1}^{k} \sigma_j^W}{\sum_{j=1}^{k} d_j^B} \tag{5.33}$$

Le paramètre $\beta \in [0{,}5\,;\,1[$ dépend des positions relatives des classes : plus les données sont dispersées, plus β doit être petit.

5.4.3.3 Apprentissage Hybride

Le mode d'apprentissage que nous utilisons est un processus hybride en deux temps : dans une première étape, les centroïdes et largeurs des FBR sont ajustés *via* une descente du gradient, puis la matrice des poids est estimée par la technique des moindres carrés. La matrice des sorties désirées, notée S, est de taille $k \times N$ et ne contient que des '0' et des '1', avec un '1' par colonne, dont l'indice correspond à la classe-cible.

Dans un premier temps, l'ajustement des paramètres $\{C_i,\sigma_i\}_{i=\{1\ldots r\}}$ de la couche cachée et des poids W est effectué de manière à minimiser la fonction de coût suivante :

$$E = \sum_{l=1}^{N} E^l = \frac{1}{2}\sum_{l=1}^{N}(S_{\cdot l} - Y_{\cdot l})^T(S_{\cdot l} - Y_{\cdot l}) \tag{5.34}$$

où $Y_{\cdot l}$ (respectivement $S_{\cdot l}$) est la $l^{\text{ème}}$ colonne de la matrice des sorties obtenues Y (respectivement de la matrice des sorties désirées S). À chaque époque e et pour chaque exemple x_l, les centres et les écarts-types sont ajustés selon :

$$C_i = C_i - \xi_c \Delta C_i^l \tag{5.35}$$

$$\sigma_i = \sigma_i - \xi_\sigma \Delta \sigma_i^l \tag{5.36}$$

où les variations ΔC_i^l et $\Delta\sigma_i^l$, respectivement associées aux taux d'apprentissage ξ_c et ξ_σ, peuvent être calculés comme suit :

$$\forall\, m \in \{1\ldots n\},$$

$$\Delta C_i^l(m) \;= \frac{\partial E^l}{\partial C_{i\,m}} = -\frac{(x_{l\,m} - C_{i\,m})}{\sigma_i^2(\sum_{i=1}^{r} R_{i\,l})^2}\, R_{i\,l}\,(\sum_{i=1}^{r} R_{i\,l} - R_{i\,l}) \sum_{j=1}^{k} W_{j\,i}(S_{j\,l} - y_{j\,l}) \tag{5.37}$$

et

$$\Delta\sigma_i = \frac{\partial E^l}{\partial \sigma_i} = -\frac{\sum_{m=1}^{n}(x_{l\,m} - C_{i\,m})^2}{\sigma_i^3\,(\sum_{i=1}^{r} R_{i\,l})^2}\, R_{i\,l}\,(\sum_{i=1}^{r} R_{i\,l} - R_{i\,l}) \sum_{j=1}^{k} W_{j\,i}(S_{j\,l} - y_{j\,l}) \tag{5.38}$$

Dans un second temps, une fois que les paramètres $\{C_i,\sigma_i\}$ ont été fixés, la matrice de poids W est ajustée en appliquant la méthode des moindres carrés sur les couples $(Y_{\cdot l},S_{\cdot l})_{l=\{1\ldots N\}}$.

5.4.3.4 Classification

Lorsqu'un visage-requête T doit être reconnu, on le projette dans l'espace défini par l'ADB selon la formule (5.1). On obtient ainsi sa signature $T^{(Q,P)}$, que l'on transforme en un vecteur t

par concaténation de ses lignes. Pour chaque unité FBR, on calcule alors la fonction d'activation $R_i(t)$ associée :

$$\forall i \in \{1,\ldots,r\},\ R_i(t) = \frac{e^{-\frac{1}{2\sigma_i^2}(t-C_i)^T(t-C_i)}}{\sum_{j=1}^{r} e^{-\frac{1}{2\sigma_j^2}(t-C_j)^T(t-C_j)}} \tag{5.39}$$

On en déduit sa sortie associée $y(t) = [y_1(t),\ \ldots,y_k(t)]^T$ défini comme suit :

$$\forall j \in \{1,\ldots,k\},\ y_j(t) = W_0 + \sum_{i=1}^{r} W_{j\,i} R_i(t) \tag{5.40}$$

Nous avons vu que chacun des éléments $y_j(t)$ est une estimation de la probabilité *a posteriori* d'appartenance à la classe associée. Par conséquent, si l'on travaille *en monde fermé*, c'est-à-dire que l'on sait que le visage-requête correspond à une personne enregistrée dans la base d'apprentissage, on choisit de lui assigner l'identité la plus probable : si l'on note j^* l'indice de l'élément maximal du vecteur $y(t)$, alors le visage-requête T est assigné à la classe Ω_{j^*}. En revanche, si l'application est *en monde ouvert*, il nous faudra appliquer une phase préliminaire de préfiltrage des visages, nous permettant de décider si un visage appartient ou non à la base d'apprentissage. Si le visage-requête est classé comme n'appartenant pas à la base d'apprentissage, on peut simplement le rejeter, ou bien l'utiliser pour augmenter la base d'apprentissage, selon les applications visées. Si par contre le visage est enregistré dans la base d'apprentissage, on se ramène alors à une reconnaissance en monde fermé et l'on applique l'algorithme de classification détaillé ci-avant.

5.4.3.5 Préfiltrage des visages

Le préfiltrage des visages se fait par l'étude de leurs vecteurs de sortie $y(t)$ associés. Notons j_l^1 la classe la plus probable pour le $l^{\text{ème}}$ visage. Nous décidons que la signature t correspond à une personne connue (en l'occurrence la classe j^1) si et seulement si :

$$y_{j^1}(t) \simeq \mathbb{P}[j^1/t] > \tau \tag{5.41}$$

où le seuil τ peut être déterminé à l'aide de l'utilisation d'une base de validation contenant à la fois des visages enregistrés et des visages inconnus. Celui-ci doit fournir un bon compromis entre taux de *fausses alarmes* et taux de *faux rejets* (*cf.* section 1.7). Un *faux rejet* consiste à décider qu'un visage donné n'est pas enregistré dans la base d'apprentissage, alors qu'il l'est. Une *fausse alarme*, quant à elle, consiste à décider qu'un visage est connu de la base de connaissance alors que cela n'est pas le cas. Une telle stratégie vise à déterminer une valeur du paramètre τ qui soit la plus adaptée possible à la complexité de la base d'apprentissage (dispersion des données, hétérogénéité des bases, etc.).

5.4.4 Évaluation de la méthode proposée

Dans cette section, nous évaluons les performances de la technique proposée pour l'identification en monde ouvert et fermé. En monde ouvert, les taux de *faux rejets*, de *fausses alarmes* et les *taux de reconnaissance* (*cf.* section 1.7) sont comparés à ceux obtenus par une technique basée sur les *eigenfaces* pour l'extraction de signatures, et les RFBRN pour leur classification (ACP+RFBRN). En monde fermé, les taux de reconnaissance sont comparés à cette dernière méthode (ACP+RFBRN), ainsi qu'aux techniques d'ADB, d'ADL2Do, de l'ACP2D, des *fisherfaces* et des des *eigenfaces* utilisées conjointement avec une distance D_{L_2} au plus proche voisin.

5.4.4.1 Protocole expérimental

Les expérimentations sont menées sur une sous-base de la base Asian Face Image Database PF01 (*cf.* annexe A), contenant des vues de 75 personnes, dont aucune ne porte de lunettes et sous des conditions d'illumination neutres. Les images de visages sont normalisées comme décrit en annexe C, puis redimensionnées à une taille suffisante de 75×65 pixels. La base d'apprentissage contient quatre vues par personne pour 60 des 75 personnes de la base, avec une expression faciale neutre et des poses proches de la pose frontale (voir figure 5.19-a). Cette base d'apprentissage sert également de base de connaissance.

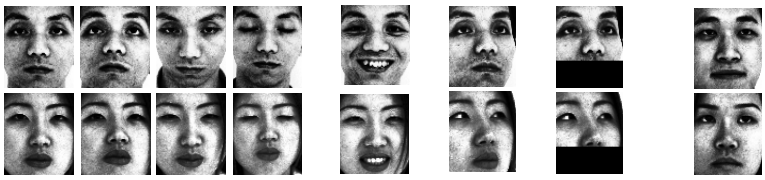


FIG. 5.19 – *Bases utilisées pour la première expérimentation : (a) base d'apprentissage, (b) base de test « expression » , (c) base de test « pose » , (d) base de test « occultation » , (e) base de test « inconnus », contenant des visages de personnes non enregistrées dans la base d'apprentissage.*

On considère quatre bases de test visant à tester l'approche en monde fermé ou ouvert. Tandis que les trois premières bases (voir figure 5.19-b-d) ne contiennent que des vues des 60 personnes connues, la quatrième base (*cf.* figure 5.19-e) contient des vues de 15 personnes non enregistrées dans la base d'apprentissage.

Chacune des trois premières bases de test est constituée d'une vue par personne, pour chacune des 60 personnes de la base d'apprentissage, dans des conditions différentes des conditions d'apprentissage. Tandis que les deux premières bases représentent respectivement des changements d'expression faciale (base « expression ») et de pose (base « pose »), la troisième base (« occultation ») simule une occultation partielle de la partie basse du visage (port d'une écharpe par exemple) par l'ajout d'une bande de pixels noirs depuis le bas du visage jusqu'à 25 pixels de haut (sur 75 pixels de hauteur totale). Notons que les techniques statistiques globales sont réputées peu tolérantes à des modifications importantes de cette région faciale [GSC01] ; nous cherchons donc à caractériser le comportement de la technique proposée dans ce contexte. Ces trois bases sont confrontées à la base d'apprentissage dans le but d'évaluer les taux de faux rejets, ainsi que les taux de reconnaissance.

La quatrième base contient une image pour chacune des 15 personnes non enregistrées. Notons que ces images sont tirées de la base PF01 et que les conditions de prise de vue (illumination, pose, etc.) sont les mêmes que dans la base d'apprentissage. Ces bases serviront à l'évaluation des taux de fausses alarmes.

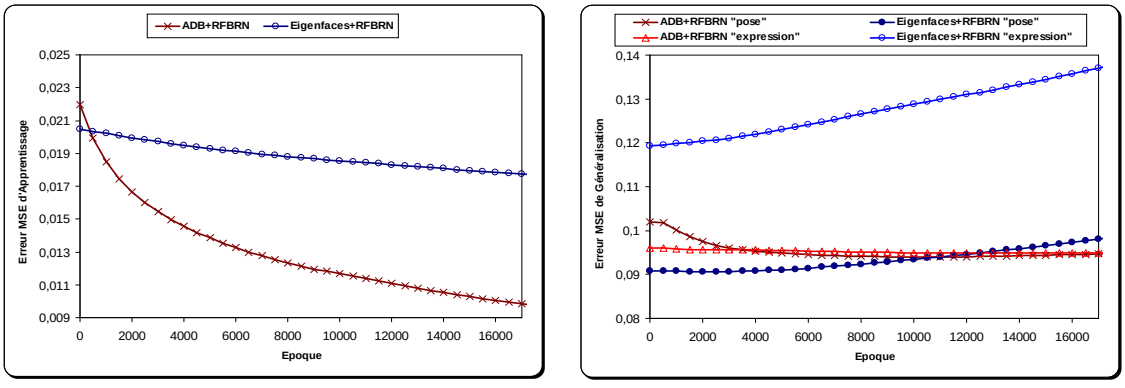
5.4.4.2 Construction du modèle

À partir de la base d'apprentissage, on construit trois modèles d'extraction de signature par projection statistique : un modèle d'*eigenfaces*, un de *fisherfaces*, et un d'ADB. Les signatures (vecteurs de coefficients) obtenus par le biais des *eigenfaces* servent directement à l'apprentissage des RFBRN, tandis que pour l'ADB les matrices-signatures sont préalablement vectorisées (par concaténation de leurs lignes de pixels par exemple). Les signatures issues des *fisherfaces* sont directement comparées à l'aide d'une mesure de dissimilarité. Les nombres de vecteurs propres

retenus pour chacun des trois modèles sont respectivement $g_{ADB} = 15$, $g_{ACP} = 150$ et $g_{ADL} = 59$ pour l'ADB, les *eigenfaces* et les *fisherfaces* (les *fisherfaces* sont construit depuis un nombre suffisant de 150 *eigenfaces*).

Concernant la construction des RFBRN (pour les *eigenfaces* et l'ADB), des expérimentations préliminaires ont mis en lumière le fait que c'est la seconde stratégie d'initialisation présentée en section 5.4.2.3 (utilisant un FBR par classe) qui offre le meilleur compromis entre stabilité de la convergence et performances du système. Le réajustement des variances est effectué avec $\beta = 0{,}7$. Les taux d'apprentissage utilisés pour les RFBRN sont $\xi_c = \frac{\|C\|_2}{n}$ et $\xi_\sigma = \frac{\|\sigma\|_2}{n}$; ils sont remis à jour toutes les 1000 époques. Le seuil de rejet (5.41) est fixé à $\tau = 0{,}25$. Cette valeur minimise sur la base considérée le taux de faux rejets, tout en garantissant un nombre très faible de fausses alarmes.

Intéressons-nous à l'apprentissage des RFBRN pour les *eigenfaces* et l'ADB. Les erreurs moyennes quadratiques des deux approches (en anglais *Mean Square Error*, notée ici MSE), mesurant l'écart entre les sorties obtenues et désirées sur la base d'apprentissage, sont comparées en figure 5.20-a. Nous pouvons remarquer que, si les MSE des deux méthodes sont initialement comparables et décroissent au fil de l'apprentissage, le MSE des *eigenfaces*+RFBRN diminue bien plus lentement que le MSE de l'ADB+RFBRN . Cela pourrait nous laisser à penser que l'apprentissage des RFBRN pour les *eigenfaces* est trop lent, mais cela est contredit par la figure 5.20-b. Celle-ci donne les MSE calculées sur les deux premières bases de test, en fonction de l'époque d'apprentissage. On peut noter que, tandis que les MSE de l'ADB+RFBRN diminuent au fil de l'apprentissage, les MSE des *eigenfaces*+RFBRN augmentent sans interruption. Pourtant, les taux d'apprentissage ont été choisis de manière à garantir les meilleurs taux de reconnaissance. Tout se passe donc comme si l'ADB+RFBRN avait un meilleur pouvoir de généralisation que les *eigenfaces*+RFBRN. Les taux de reconnaissance qui seront donnés en section 5.4.4.4 viendront confirmer cette hypothèse. Ces mêmes résultats permettront également de mettre en lumière que, pour les *eigenfaces*+RFBRN, les taux de reconnaissance augmentent avec le nombre d'époques malgré l'augmentation des MSE et que par conséquent l'apprentissage des RFBRN sur les signatures des *eigenfaces* s'est déroulé correctement. Pour les deux techniques, on considère les apprentissages terminés au bout de 8000 époques.



(a) **(b)**

FIG. 5.20 – *Erreurs moyennes quadratiques (MSE) des eigenfaces+RFBRN et de l'ADB+RFBRN, calculés depuis (a) la base d'apprentissage et (b) les deux premières bases de test, en fonction du nombre d'époques de l'apprentissage des RFBRN.*

5.4.4.3 Préfiltrage des visages

Dans cette section, nous cherchons à caractériser l'efficacité de la règle de filtrage des visages (selon qu'ils appartiennent ou non à la base d'apprentissage). Intéressons-nous tout d'abord au taux de faux rejets. Nous utilisons pour cela les trois premières bases (*cf.* figure 5.19-b–d), contenant exclusivement des personnes enregistrées. Pour la technique d'*eigenfaces*+RFBRN comme pour celle d'ADB+RFBRN, le nombre de rejets sur ces trois bases (donc le taux de faux rejets) est nul, après un nombre d'époques (8000) suffisant pour un apprentissage efficace des RFBRN. Et ceci même sur la base « occultation ». Ce dernier point est particulièrement important. En effet, cela signifie qu'il ne suffit pas d'occulter la partie basse de son visage pour être classé comme inconnu, si l'on est enregistré.

Le taux de fausses alarmes est évalué sur la base des personnes inconnues (*cf.* figure 5.19-e). La figure 5.21 montre que, malgré des conditions très similaires à celles de la base d'apprentissage, le taux de fausses alarmes est très faible pour les deux approches évaluées : on enregistre 1 fausse alarme pour l'ADB+RFBRN, et 2 fausses alarmes pour les *eigenfaces*+RFBRN (différence de performance non significative).

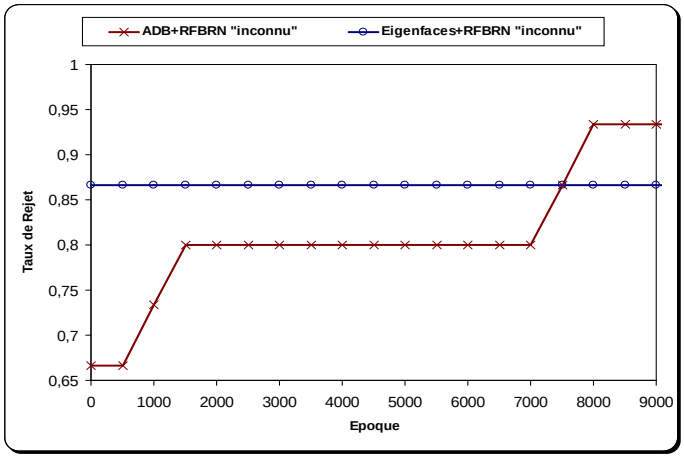


FIG. 5.21 – *Evolution des taux de rejet des deux méthodes évaluées, calculés sur la base des visages non enregistrés, en fonction du nombre d'époques.*

5.4.4.4 Évaluation des taux de reconnaissance en monde fermé

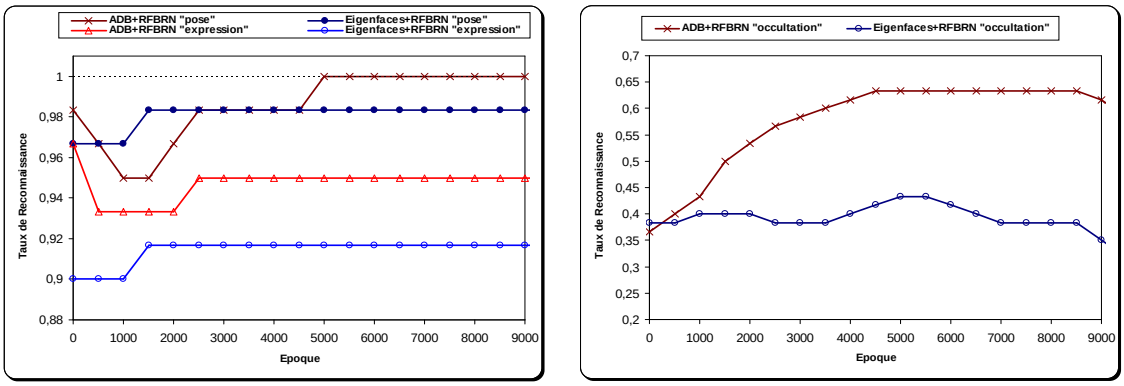
Une fois le processus de préfiltrage des visages effectués, on considère que l'on a rejeté les visages n'appartenant pas à la base. On se ramène alors à des processus d'identification en monde fermé, dont nous évaluons les performances en mesurant les taux de reconnaissance obtenus sur les trois bases de test des personnes enregistrées. Les taux de reconnaissance comparés des *eigenfaces*, des *fisherfaces*, de l'ACP2D, de l'ADL2DoL, de l'ADL2DoC et de l'ADB utilisant la distance Euclidienne au plus proche voisin sont donnés en table 5.4. Celle-ci nous servira de point de comparaison avec les performances observées en utilisant les RFBRN.

La figure 5.22 donne les taux de reconnaissance comparés des techniques d'ADB+RFBRN et d'*eigenfaces*+RFBRN, en fonction du nombre d'époques de l'apprentissage. On note que, malgré les MSE croissants pour l'apprentissage, les taux de reconnaissance des *eigenfaces*+RFBRN s'améliorent quand le nombre d'époques augmente. On remarque en comparant cette figure avec la table 5.4 que, si l'utilisation des RFBRN améliore systématiquement les taux de reconnais-

	Eigenfaces	Fisherfaces	ACP2D	ADL2DoC	ADL2DoL	ADB
« pose »	96,7%	98,3%	96,7%	96,7%	98,3%	100 %
« expression »	86,7%	95%	93,3%	95%	96,7%	96,7%
« occultation »	13,3%	50%	55%	70%	53,3%	70%

TAB. 5.4 – *Taux de reconnaissance comparés des eigenfaces, des fisherfaces, de l'ADCP2D, de l'ADL2DoC, de l'ADL2DoL et de l'ADB, avec une distance Euclidienne au plus proche voisin.*

sance des *eigenfaces*, il n'en est pas toujours de même pour l'ADB, notamment sur les bases « expression » et « occultation » (les baisses sont respectivement de 1 et 3 visages sur 60 par rapport à la distance D_{L_2}). Cependant, la technique d'ADB+RFBRN reste supérieure aux techniques des *eigenfaces*+RFBRN, *eigenfaces*+D_{L_2}, *fisherfaces*+D_{L_2} et , ACP2D+D_{L_2}, sur les trois bases de test considérées. Il faut également noter que l'utilisation des RFBRN à la place de la distance D_{L_2} au plus proche voisin nous permet de réduire de manière très importante le coût de la classification, ce qui constitue un avantage important à mettre au crédit des RFBRN. Puisque de plus les RFBRN nous permettent de travailler de manière efficace en monde ouvert, nous préconisons d'utiliser les RFBRN pour la classification des signatures issues de l'ADB, plutôt qu'une mesure de dissimilarité au plus proche voisin.

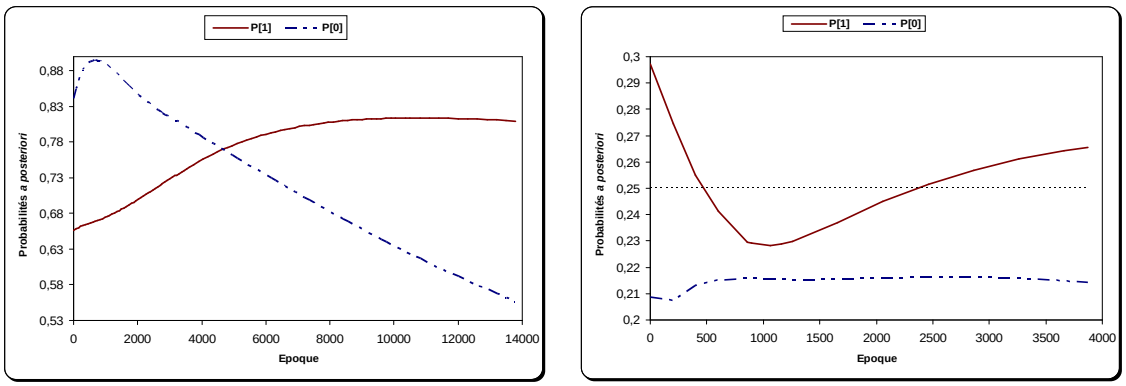


(a) **(b)**

FIG. 5.22 – *Évolution des taux de reconnaissance des eigenfaces+RFBRN et de l'ADB+RFBRN, calculés (a) depuis les bases de test « expression » et « pose » et (b) à partir de la base de test « occultation », en fonction du nombre d'époques de l'apprentissage des RFBRN.*

Intéressons-nous maintenant au cas particulier des visages provenant de la base de test « pose » et qui, à un moment ou à un autre du processus d'apprentissage, sont mal classés ou faussement rejetés par l'ADB+RFBRN. Au fil de la phase d'apprentissage, trois erreurs sont constatées, mais après 4800 époques toutes ont convergé vers leur classe d'appartenance et le taux de reconnaissance atteint 100%. Notons « $\mathbb{P}[1]$ » la probabilité d'appartenance à la classe-cible et « $\mathbb{P}[0]$ » la probabilité d'appartenir à la mauvaise classe la plus probable. L'erreur illustrée en figure 5.23-a correspond à une erreur de classification : en d'autres termes, $\mathbb{P}[0]>\mathbb{P}[1]$. Cette erreur, déjà présente à l'initialisation, est résolue après 4800 époques, lorsque $\mathbb{P}[1]$ devient supérieure à $\mathbb{P}[0]$ (tout en restant supérieure à $\tau = 0{,}25$). L'erreur illustrée en figure 5.23-b correspond à

un faux rejet, dû au fait que $\mathbb{P}[1]$, bien que supérieure aux probabilités des autres classes, est inférieure à τ entre les époques 500 et 2450. La troisième erreur constatée (non illustrée dans la figure 5.23) correspond à un cas semblable de faux rejet, survenant entre les époques 500 et 1950. Dans ces deux derniers cas, la probabilité $\mathbb{P}[1]$ devient supérieure à $\tau = 0{,}25$ après un nombre suffisant d'époques, ce qui montre la capacité de généralisation de la technique proposée.



(a) **(b)**

FIG. 5.23 – *Probabilités d'appartenance à la classe-cible (« $\mathbb{P}[1]$ »), et à la mauvaise classe dont la probabilité associée est la plus élevée (« $\mathbb{P}[0]$ »), en fonction du nombre d'époques, pour (a) l'erreur de classification et (b) l'un des deux faux rejets.*

On peut noter que le taux de reconnaissance calculé sur la base « occultation » (70% pour l'ADB+RFBRN) est très inférieur à celui des autres bases. L'occultation fait donc chuter les taux de reconnaissance. Si l'on étudie plus en détail les résultats de classification, on s'aperçoit que, dans tous les exemples de mauvaise classification, la classe désirée se voit attribuer une probabilité d'appartenance supérieure à $\tau = 0{,}25$, mais que c'est une autre classe qui est la plus probable, parfois avec très peu d'écart. Cela met en évidence que, parfois, l'ambiguïté entre deux classes d'apprentissage est trop importante. Pour pallier ce problème, on peut envisager de mettre en œuvre une technique basée sur un ajout incrémental de nouvelles unités FBR au cours du processus d'apprentissage, aux endroits où cela est nécessaire. On pourrait par exemple envisager d'ajouter un FBR au centre du segment reliant les centres de deux classes, si celles-ci se chevauchent de manière trop importante, et que ce chevauchement ne diminue pas de manière satisfaisante au cours de l'apprentissage (selon une mesure à définir).

5.4.5 Conclusion

Dans cette section, nous avons présenté un classifieur efficace dans le cadre de l'identification de visages en monde ouvert : il permet dans un premier temps de rejeter les visages de personnes non enregistrées dans la base d'apprentissage, et de ne garder que les visages connus pour la seconde phase de classification en monde fermé. Cette seconde phase consiste à assigner à chaque visage-requête une classe d'appartenance parmi celles représentées dans la base de connaissance.

La technique proposée consiste à extraire les signatures des visages à l'aide de l'ADB puis, à partir des signatures ainsi obtenues, d'entraîner un Réseau de Fonctions à Base Radiale Normalisé pour la classification. L'amélioration apportée est double. Premièrement, l'utilisation de RFBRN diminue drastiquement le coût de la phase de classification par rapport à une mesure de distance au plus proche voisin, et permet d'aborder des applications en monde ouvert. Deuxièmement, nous avons montré par le biais d'expérimentations que l'approche ainsi proposée fournit

de meilleurs taux d'identification que les techniques usuelles de projection statistique, tout en fournissant des taux de fausses alertes et de faux rejets très faibles.

5.5 Conclusion

Dans ce chapitre, nous avons présenté une technique novatrice de reconnaissance de visages. Nous avons tout d'abord mis au point une stratégie efficace d'extraction de signature : il s'agit de l'ADB, qui combine efficacement les avantages de chacune des deux versions issues de l'ADL2Do, et donne de meilleurs résultats de classification que la plupart des autres techniques de projection statistique. Afin d'accroître la robustesse à des occultations partielles, nous avons prouvé l'efficacité d'une utilisation modulaire de l'ADB. Ces techniques d'extraction de signatures sont combinées avec la distance euclidienne au plus proche voisin, dont nous avons montré qu'il s'agit de la mesure de dissimilarité (parmi les mesures de Minkowski et de Minkowski fractionnaires) la plus performante dans la plupart des cas pour la classification des signatures extraites. Nous avons montré les excellents performances des approches proposées, en comparaison avec celles de l'état de l'art.

Puis, dans le but de réduire le coût calculatoire de la phase de classification et de définir des règles nous permettant d'aborder des problèmes d'identification en monde ouvert, nous avons introduit l'utilisation de Réseaux de Fonctions à Base Radiale Normalisés pour la classification des signatures issues de l'ADB. Nous avons montré expérimentalement les très bonnes performances de l'approche proposée pour l'identification de visages en monde ouvert.

Conclusion et perspectives

Cette thèse s'inscrit dans le contexte de l'identification automatique de visages dans des images numériques par le biais de techniques de projection statistique. Nous avons introduit des méthodes d'extraction de signatures mettant en œuvre une Analyse Discriminante de manière à ce que l'information bidimensionnelle provenant des images soit prise en compte. L'utilisation conjointe de Réseaux de Fonctions à Base Radiale Normalisés permet d'obtenir d'excellentes performances, tant dans un contexte de monde fermé qu'ouvert. Dans cette section, nous rappelons dans un premier temps les apports de cette thèse, avant de discuter des avantages et des limitations des approches introduites. Enfin, nous proposons des directions de recherche se plaçant dans la continuité de nos travaux.

Synthèse

Depuis les trente dernières années, la reconnaissance automatique de visages est un domaine de recherche très actif, et de nombreuses méthodes ont été proposées dans ce contexte. Parmi les techniques les plus efficaces, on compte les méthodes de projection statistique. Celles-ci trouvent leurs origines dans l'analyse des propriétés statistiques des images de visages : du fait de la redondance de l'information et de la présence de bruit dans celles-ci, on peut utiliser un espace réduit de projection $\mathcal{F}$ pour mener à bien la classification. Plaçons-nous dans un contexte d'identification en monde fermé. Les visages sont projetés dans l'espace $\mathcal{F}$ (le mode de projection dépendant de la nature de ce dernier). Les coordonnées ainsi obtenues définissent des signatures. Les signatures des visages à reconnaître sont comparées à celles d'une base de connaissance, le plus souvent par le biais d'une mesure de dissimilarité. Une stratégie d'affectation (généralement au plus proche voisin) permet d'assigner à chaque visage à reconnaître une identité parmi celles enregistrées dans la base d'apprentissage. Les techniques de projection statistique visent à définir un espace de projection doté de bonnes propriétés, en termes de représentativité des données d'entrée ou de séparation des classes. Le calcul de $\mathcal{F}$ est généralement mené à l'aide d'une technique d'analyse de données multidimensionnelles, telle que l'Analyse en Composantes Principales (ACP) ou l'Analyse Discriminante Linéaire (ADL). Cette analyse est traditionnellement menée sur des vecteurs-visages, définis à partir des images de la base d'apprentissage par concaténation de leurs lignes (ou colonnes) de pixels. Cette modélisation des données d'entrée est qualifiée d'*unidimensionnelle* (1D).

Ces vecteurs étant généralement de très grande taille comparé à leur nombre, on est confronté à un problème de sous-représentation des données qui pose un certain nombre de difficultés. Premièrement, l'estimation des paramètres des modèles est instable et coûteuse. De plus, *le problème de la singularité* empêche une application directe de l'ADL, qui est pourtant une technique conçue pour la classification. Afin de contourner ce problème, de nombreuses variantes de l'ADL ont été introduites. La modification peut porter sur les données d'entrée (*fisherfaces*), le mode de mise

en œuvre de l'ADL (ADL sous-optimale) ou le critère à maximiser (ADL modifiée). Les performances de ces techniques sont très bonnes, en comparaison avec les autres méthodes de l'état de l'art. Cependant, leur construction repose la plupart du temps sur des estimations coûteuses et/ou instables numériquement et leur efficacité peut varier en fonction des bases de visages considérées. Certaines reposent sur un ajustement difficile et coûteux de paramètres supplémentaires introduits par la modification apportée.

L'ADL Bidimensionnelle Orientée

Afin de pallier ces inconvénients, nous avons introduit une technique globale baptisée Analyse Discriminante Linéaire Bidimensionnelle Orientée (ADL2Do). Celle-ci se décline en deux versions. La première (ADL2DoL) consiste à appliquer une ADL sur les lignes des images, l'autre (ADL2DoC) sur leurs colonnes. Ces modes de représentations des données sont qualifiés de *bidimensionnels orientés* (2Do). Nous avons montré qu'utiliser une modélisation 2Do des visages revient à augmenter artificiellement le nombre d'exemples disponibles, tout en réduisant leur taille (par rapport à une représentation 1D). Ainsi, l'ADL2Do contourne le problème de la singularité sans ajout de paramètres supplémentaires. De plus, le coût et l'instabilité numérique sont réduits lors de la construction du modèle. Cependant, la taille des signatures est plus importante que pour les techniques 1D, ce qui rend la phase de classification plus coûteuse. Nous avons montré sur différentes bases de visages internationales les très bonnes performances de l'ADL2DoL et de l'ADL2DoC ainsi que leur complémentarité en termes de résultats de classification. Une combinaison efficace de ces deux techniques peut donc permettre de définir une approche plus performante que chacune d'elles prise séparément.

L'Analyse Discriminante Bilinéaire

C'est dans cette optique que nous avons proposé la technique globale nommée Analyse Discriminante Bilinéaire (ADB). La projection linéaire utilisée dans le contexte de l'ADL est remplacée par une projection bilinéaire. Au lieu de rechercher l'espace de projection séparant au mieux par projection linéaire les différentes classes, on cherche le couple de matrices de projection qui, par projection bilinéaire, permet de classer au mieux les données. On considère ce modèle comme *bidimensionnel*, car il tire avantage des deux représentations 2Do des données. La taille des signatures est considérablement réduite par rapport à l'ADL2Do. La fonction objectif à maximiser étant biquadratique, il n'existe pas de solution analytique à ce problème d'optimisation. C'est pourquoi nous avons proposé un mode de mise en œuvre itératif. Nous avons mis en évidence les excellentes performances de celui-ci par le biais d'expérimentations rigoureuses menées sur diverses bases de visages. Celles-ci montrent clairement que l'ADB est un mode de combinaison efficace des deux versions issues de l'ADL2Do. Les excellentes performances observées placent l'ADB parmi les meilleures techniques de projection statistique globales pour l'identification en monde fermé. Nous avons aussi montré en section 2.3.2 que les techniques dites *hybrides*, au sens où elles allient les avantages d'une représentation globale et locale des visages, sont souvent plus efficaces et tolérantes à différents types de variation (notamment l'expression faciale) que les approches purement globales.

L'Analyse Discriminante Bilinéaire Modulaire

C'est pourquoi nous avons introduit la méthode d'Analyse Discriminante Bilinéaire Modulaire (ADBM), qui repose sur l'utilisation conjointe de trois experts entraînés indépendamment sur des régions faciales différentes. L'un d'entre eux est entraîné sur la globalité du visage, les

deux autres sur des régions faciales localisées dans la région supérieure du visage. Ces dernières sont choisies de manière à être complémentaires vis-à-vis des différentes sources de variabilité et ainsi à garantir de bonnes performances dans la plupart des cas. Différents modes de combinaison de ces experts ont été évalués. Le premier, nommé *agrégation d'experts*, consiste à combiner les résultats de classification obtenus par les trois experts. Le second mode de combinaison est baptisé *fusion d'experts* et est basé sur une fusion des signatures fournies par les trois experts pour définir une méta-signature, qui sert à la classification des visages. Nous avons montré que l'ADBM, et surtout la fusion d'experts, est plus performante que chacun des experts considérés séparément et qu'une technique modulaire basée sur une ACP 1D [PMS94]. La tolérance à des changements d'expression faciale est notamment améliorée. L'ADBM, comme l'ADB et l'ADL2Do, utilise une classification basée sur la mesure d'une distance au plus proche voisin. Cette règle d'affectation présente le désavantage d'être coûteuse, et potentiellement influencée par des observations aberrantes.

L'utilisation de Réseaux de Fonctions à Base Radiale Normalisés

Afin de corriger ces inconvénients nous proposons, dans le contexte de la classification des signatures issues de l'ADB, de remplacer la distance au plus proche voisin par un Réseau de Neurones à Base Radiale Normalisé (RFBRN). Cette technique permet de modéliser les classes de signatures avec un faible nombre de paramètres, ce qui rend la phase de classification beaucoup moins coûteuse en termes de temps de calcul. Les RFBRN fournissent en sortie des mesures normalisées que l'on peut interpréter comme des estimations des probabilités *a posteriori* d'appartenance à chacune des classes. Afin d'appliquer l'approche proposée à un contexte de monde ouvert, nous dérivons de ces estimations des règles de décision simples et permettant un filtrage efficace des visages non enregistrés dans la base. Nous avons mené une évaluation rigoureuse de la technique proposée. Cette étude a permis de mettre en lumière l'efficacité de celle-ci en monde ouvert : sur les bases utilisées, le taux de faux rejet est nul et le taux de fausses alarmes, très faible. En monde fermé, les taux de reconnaissance obtenus sont meilleurs qu'avec la plupart des techniques de projection statistique usuelles (utilisant une distance au plus proche voisin). En monde fermé et ouvert, les performances du système sont supérieures à celles des *eigenfaces*, utilisées conjointement avec un RFBRN. Nos résultats mettent également en lumière la très bonne tolérance du système proposé vis-à-vis de changements d'expression faciale, d'angle de prise de vue et d'occultations partielles.

Discussion

Nous avons montré dans cette thèse les excellentes performances des systèmes proposés, notamment de l'ADB avec un schéma de classification basé sur simple mesure de dissimilarité ou sur l'utilisation de RFBRN. Ces performances ont été évaluées sur quelques-unes des principales bases de visages internationales, selon des protocoles expérimentaux rigoureux. Cependant, toutes les études tendent à prouver que le passage à une application réelle engendre une baisse des performances de la grande majorité des systèmes de reconnaissance de visages. Nous avons mené des expérimentations sur des images de visages que nous avons collectées à l'aide d'une *webcam* et constaté que l'ADB n'échappe pas à cette règle. Cela est notamment dû au fait que de nombreuses sources de variabilité peuvent cohabiter dans les images de visages à reconnaître, ce qui n'est généralement pas le cas dans les bases de visages utilisées pour l'évaluation. De plus l'ADB, à l'instar de la plupart des techniques de projection statistique, nécessite la mise en œuvre

d'une phase préliminaire de détection et de segmentation des visages dans l'image. Dans l'expérimentation que nous avons menée, les visages étaient simplement détectés et segmentés dans l'image, sans localisation précise de leurs caractéristiques faciales. Les images ainsi segmentées étaient donc sujettes à des changements d'échelle du visage et à la prise en compte d'une partie du fond –complexe dans les conditions d'évaluation retenue– de l'image. Nous avons constaté que ces artefacts étaient responsables d'une importante partie des mauvaises classifications. Les erreurs restantes étaient essentiellement dues à des changements dans les conditions d'illumination entre deux prises de vue. Les enseignements que nous avons tirés de cette expérimentation, ainsi que les évaluations que nous avons menées sur les bases de visages, peuvent nous aider à définir les grandes lignes d'un système efficace dans le contexte d'une application réelle. C'est l'objet de la section suivante.

Perspectives

Dans cette section, nous nous intéressons à l'application de nos travaux dans le contexte d'applications réelles. Le choix des techniques à mettre en œuvre est directement lié aux contraintes de l'application visée. Considérons dans un premier temps l'exemple de la biométrie dans un cadre familial (*cf.* section 1.3), avant de généraliser notre propos à d'autres familles d'applications.

Biométrie dans un cadre familial

Plaçons-nous dans le contexte d'une application de biométrie dans un cadre familial. Il s'agit d'une tâche d'identification en monde fermé. Le nombre de personnes enregistrées étant petit, on peut considérer que le coût de la règle d'affectation basée sur une mesure de distance au plus proche voisin reste raisonnable. Dans le but de favoriser la convivialité du système en diminuant le coût de construction du modèle, nous proposons d'appliquer la technique d'ADBM plutôt que celle basée sur l'utilisation conjointe de l'ADB et des RFBRN.

Les utilisateurs sont amenés à interagir avec le système lors de deux grandes étapes : l'enregistrement des personnes à reconnaître et la phase de reconnaissance proprement dite. Entre-temps, le système doit avoir construit un modèle des visages enregistrés permettant de mener à bien leur classification. La suite de cette section constitue une discussion autour des stratégies à adopter lors de ces trois grandes phases (enregistrement, construction du modèle, reconnaissance) dans le contexte de la biométrie dans un cadre familial ou de tout autre application présentant les mêmes caractéristiques.

Enregistrement des personnes à reconnaître

Lors de la phase d'enregistrement, chaque utilisateur du système se place devant une *webcam*, de manière à ce que le système enregistre des vues de son visage. On demande préalablement à chaque utilisateur de s'identifier, de manière à pouvoir étiqueter chaque image collectée par son identité. La base d'apprentissage est constituée de l'ensemble de ces images.

Les expérimentations que nous avons menées en section 5.2.5.3 sur la base AR nous ont permis de mettre en évidence l'influence du contenu de la base d'apprentissage sur les performances de l'ADB. En particulier, les images de visages collectées dans des conditions difficiles d'éclairage sont plus aisément reconnues si la base d'apprentissage contient des vues prises dans des conditions similaires ou qu'au moins cette dernière contient des variations d'illumination. Afin de garantir les meilleures performances du système, il faut donc apporter une attention particulière au contenu de la base d'apprentissage. Dans cette optique, il serait souhaitable de disposer pour

chaque utilisateur d'un ensemble d'images suffisamment représentatif des différentes variations d'illumination possibles. On peut envisager pour cela d'encourager les utilisateurs à s'enregistrer dans au moins deux conditions d'illumination différentes. Le même raisonnement est applicable pour l'expression faciale : on peut par exemple demander à l'utilisateur de réciter quelques mots devant la *webcam*, de manière à enregistrer son visage dans différentes conditions.

Intéressons-nous maintenant aux principaux prétraitements à mettre en œuvre sur la base d'apprentissage. Au vu de nos résultats expérimentaux, nous pensons qu'il est indispensable d'appliquer une détection de caractéristiques faciales la plus précise possible entre les phases de détection et de normalisation du visage dans l'image. Nous pensons qu'il serait judicieux d'utiliser pour cela le détecteur de Duffner et Garcia [DG05] présenté en section 1.6.2. Comme nous l'avons montré en section 1.5.1, la détection des caractéristiques faciales permet de procéder à une classification de la pose des visages. Les vues correspondant à des poses trop différentes de la pose frontale peuvent alors être rejetées de la base d'apprentissage avant la construction du modèle. Éventuellement, si le nombre de visages restants après cette phase est insuffisant, on pourra demander à l'utilisateur de s'enregistrer à nouveau, sous un angle de prise de vue plus raisonnable.

Construction du modèle

Lors de la phase de construction du modèle, l'ADB est appliquée de manière à extraire et à stocker dans le système les signatures des visages de la base d'apprentissage. Celles-ci seront utilisées lors de la phase de reconnaissance.

Dans le cadre d'une application réelle et malgré le soin apporté à la constitution de la base d'apprentissage, il est possible les conditions de prise de vue de certaines images diffèrent significativement des autres. Une robustification (*cf.* section 3.4.3) de l'ADB pourrait alors s'avérer particulièrement utile. Notamment, afin de réduire l'influence de classes trop éloignées des autres sur la définition des espaces de projection, nous pourrions utiliser une méthodologie de repondération de la matrice de covariance inter-classe inspirée de celle proposée par Loog dans [LDHU01] (*cf.* section 3.4.3.2).

Phase de reconnaissance

Lors de la phase de reconnaissance, l'utilisateur se présente devant la *webcam* utilisée pour son enregistrement. Il n'a pas à saisir d'identifiant ; dans un but de convivialité, le système doit être capable de le reconnaître sans disposer d'aucune information *a priori* concernant son identité. Afin de garantir les meilleures performances du système, les images collectées lors de cette phase doivent être le plus similaires possibles aux images d'apprentissage. Les instructions données à l'utilisateur devront aller dans ce sens (on pourra par exemple lui déconseiller de déplacer la webcam).

On applique alors le même processus de prétraitement des visages que pour l'enregistrement. Une fois le visage correctement segmenté, on calcule sa signature associée, que l'on compare à toutes les images de la base d'apprentissage avec une mesure de distance et une stratégie d'affectation au plus proche voisin. On pourra dans ce cadre mener une étude plus approfondie des distances les plus efficaces dans le contexte de ce type d'applications réelles.

Autres applications

Dans le contexte d'autres applications telles que l'indexation de contenu personnel ou de contenu spécifique (*cf.* section 1.3), nous ne pouvons influer sur les conditions de prise de vue

des images d'apprentissage (ou de connaissance). De nombreuses sources de variabilité peuvent être présentes dans ces bases. Pour ces applications, les contraintes de convivialité sont moins importantes que dans le contexte d'applications de biométrie dans un cadre familial. On peut donc envisager de mettre en œuvre un processus d'apprentissage plus long. La stratégie de classification basée sur l'utilisation de RFBRN permettant d'approximer au mieux les distributions potentiellement complexes des bases, nous préconisons de remplacer la distance utilisée précédemment par cette méthodologie. Afin d'améliorer ses performances, on pourra alors envisager de mettre en œuvre les RFBRN de manière incrémentale, comme nous l'avions évoqué en section 5.4.4.4, p. 158. Cette méthodologie consiste en l'ajout en cours d'apprentissage de Fonctions à Base Radiale aux endroits où cela est nécessaire (zones de chevauchement entre deux classes par exemple). Cela devrait nous permettre d'obtenir un schéma de classification plus adapté aux distributions complexes des classes que l'on est susceptible de rencontrer certaines des applications considérées.

Annexe A

Les bases de visages utilisées

Comme nous l'avons vu en section 1.7, il existe une multitude de bases de visages utilisées pour l'évaluation des algorithmes de reconnaissance automatique de visages. Dans cette section, nous décrivons les bases les plus utilisées. Chacune comporte des avantages et des inconvénients. Les plus anciennes (ORL, UMIST et Yale) sont les plus documentées et sont très utiles pour comparer de nouvelles méthodes à celles de l'état de l'art. Les plus récentes (PF01, FERET, AR et PIE) contiennent plus de personnes et sont donc utiles pour des évaluations à plus grande échelle. Différents facteurs sont appliqués sur les visages (changements d'illumination, de pose, d'expression faciale, occultations et variations dans le temps). Le tableau A.1 récapitule les principales caractéristiques des bases présentées ci-après. Une liste plus complète et très détaillée, est disponible dans [Gro04].

Base	**Nombre de personnes**	**Pose**	**Illumination**	**Expression**	**Occultation**	**Nombre de sessions**
FERET	1199	1–9	1–2	2	–	2
Yale	15	1	3	6	1	1
ORL	40	–	–	–	–	–
PF01	107	8	4	5	–	1
UMIST	20	–	–	–	–	1
AR	116	1	4	4	2	2
PIE	68	13	43	4	–	1

TAB. A.1 – *Principales caractéristiques des bases de visages listées dans cette annexe. La table contient: le nombre de personnes enregistrées, le nombre de vues sous des poses et conditions d'illumination différentes, le nombre de types d'occultations représentées, ainsi que le nombre de sessions au cours desquelles des vues d'un même individu ont pu être collectées. Les cas où l'un de ces éléments n'a pas été mesuré, ou était non contrôlé durant la prise de vue, est noté « – ».*

A.1 La base FERET

La base FERET a été collectée dans le cadre du programme Facial Recognition Technology [PWHR98, PMRR00] mené par le National Institue of Standards and Technology (NIST) Américain. Il s'agit à ce jour de la plus grande base disponible gratuitement pour les chercheurs. Les images, initialement collectées depuis un appareil photographique de 35mm, ont ensuite été digitalisées. Un extrait de cette base est donnée en figure A.1. Elle contient plus de 14051 images de résolution 256×384, représentant 1199 personnes. Les images ont été collectées lors de 15 sessions entre août 1993 et juillet 1996. Pour chaque individu, on dispose de deux vues (fa et fb) avec des expressions faciales différentes (généralement, une expression neutre et un sourire). Pour 200 de ces personnes, on dispose d'une troisième image prise avec une caméra et des conditions d'illumination différentes (vues fc). Pour ces 200 individus, la base contient des vues additionnelles montrant des changements de pose en profondeur (allant du profil gauche au profil droit, *cf.* figure A.1). Pour quelques personnes, on dispose de deux autres vues collectées dans des conditions similaires à fa et fb mais à des dates différentes (vues *duplicate*). Aucune contrainte n'est imposée sur la date de la prise de vue de l'image *duplicate* I. Par contre, la vue *duplicate* II a été collectée au moins un an après la première prise de vue.

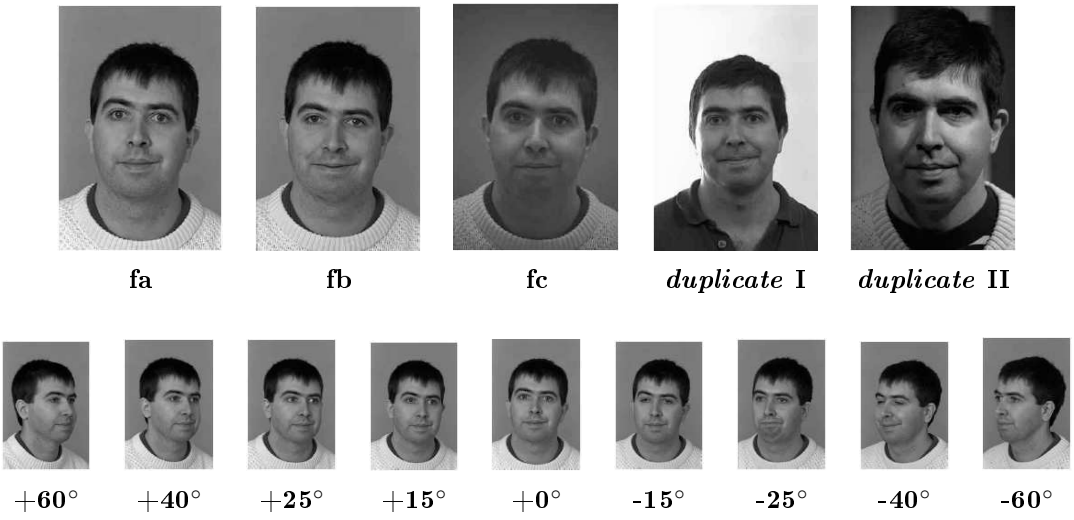


FIG. A.1 – *Extraits de la base FERET. Notons qu'il n'existe que peu de personnes pour lesquelles on dispose d'autant d'images.*

A.2 La base de Yale

La base de visages de Yale [BHK97], illustrée en figure A.2, contient 165 vues de 15 personnes (11 vues par personne). La taille des images est de 320×243 pixels. Pour chaque personne, on dispose de huit vues sous des conditions d'illumination neutres, avec : une expression faciale neutre, un sourire, le clignement d'un œil, les deux yeux fermés, une expression de surprise, une expression triste et une vue neutre avec et sans lunettes. Les trois vues restantes correspondent à des changements d'illumination : l'une avec une illumination de face, la seconde avec une illumination provenant du côté droit, la troisième venant du côté gauche. Toutes les vues d'une personne ne portant habituellement pas de lunettes (première ligne de la figure A.2) sont représentées sans lunettes, sauf une. *A contrario*, pour deux sujets (exemple en deuxième ligne de la figure A.2), on dispose de dix vues avec lunettes et d'une seule vue sans.

FIG. A.2 – *Extrait de la base de Yale. Pour chacune des 15 personnes enregistrées, on dispose de 11 vues.*

A.3 La base ORL

La base ORL [SH94] a été collectée dans le cadre d'un projet mené par un laboratoire de AT&T, basé à Cambridge, en collaboration avec l'université de Cambridge. Les prises de vue ont été menées entre avril 1992 et avril 1994. La base contient 40 personnes, chacune étant enregistrée sous 10 vues différentes (*cf.* figure A.3). Les images sont de taille 112×92 pixels. Pour quelques sujets, les images ont été collectées à des dates différentes, avec des variations dans les conditions d'illumination, les expressions faciales (expression neutre, sourire et yeux fermés) et des occultations partielles (port de lunettes ou non). Toutes les images ont été collectées sur un fond foncé. Les poses de la tête présentent quelques variations en profondeur par rapport à la pose frontale. Cette base fait partie de celles qui ont été le plus utilisées et permet de comparer facilement les performances de tout nouvel algorithme à ceux de l'état de l'art. Étant donné que les variations ne portent que sur certaines personnes et ne sont donc pas systématiques, cette base ne peut cependant pas être utilisée pour mener une analyse de sensibilité à différents facteurs.

FIG. A.3 – *Extrait de la base ORL. Pour chacune des 40 personnes enregistrées, on dispose de 10 vues avec des changements de pose, d'expression et d'illumination.*

A.4 La base PF01

La base PF01 (Postech Faces'01) [HRL04] a été développée par le laboratoire Intelligent Multimedia Laboratory (IML) de l'université de Pohang, en Corée. Elle contient des vues de 107 personnes d'origine asiatique (56 hommes et 51 femmes). Une trentaine d'entre elles portent des lunettes. Pour chaque personne, on dispose de 17 vues, comme le montre la figure A.4. Les images sont en couleurs et de taille 1280×960 pixels. Treize vues par personne sont prises dans des conditions d'illumination neutres et standardisées. Parmi ces treize vues, il y en a une qui représente une expression faciale neutre, quatre qui correspondent à des changements d'expression faciale et huit qui montrent des changements de pose en profondeur. Les quatre vues restantes, collectées sous une pose frontale et avec une expression faciale neutre, contiennent des changements d'illumination.

Cette base, disponible gratuitement sur internet sur le site d'IML http://nova.postech.ac.kr/, est très intéressante pour étudier les effets de changements de pose et d'expression faciale. Elle présente de plus l'avantage de compter de nombreuses vues pour un nombre de personnes enregistrées de l'ordre de la centaine.

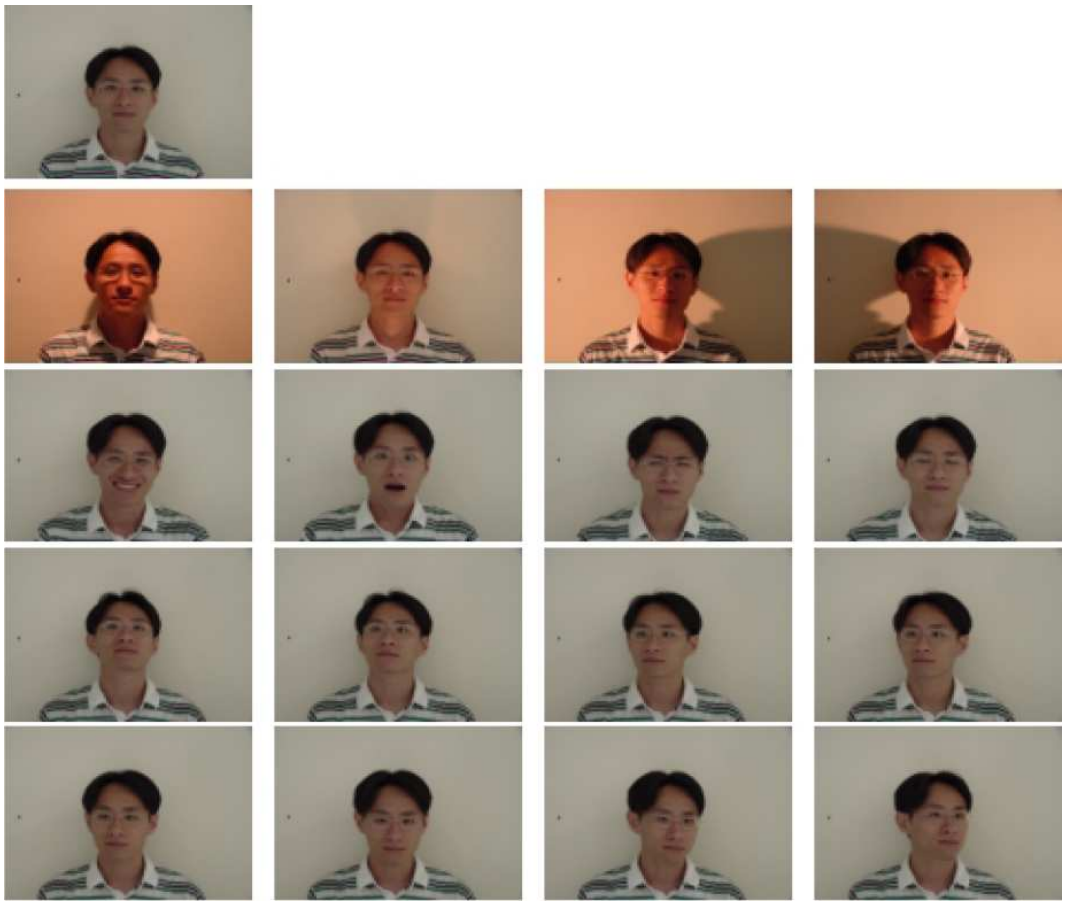


FIG. A.4 – *Extrait de la base PF01. Pour chacune des 107 personnes enregistrées, on dispose de 17 vues avec des changements de pose, d'expression faciale et de conditions d'illumination. La première image en haut à gauche représente la vue neutre. La seconde ligne montre les quatre variations d'illumination, où la direction d'illumination change : haut, bas, gauche et droit. La troisième ligne montre les changements d'expression faciale, de gauche à droite : souriant, surpris, irrité et yeux fermés. Les deux dernière lignes correspondent à des changements de pose en profondeur, allant jusqu'à 45° dans chacune des quatre direction : haut, bas, gauche, droite.*

A.5 La base UMIST

La base UMIST [GA98] contient 564 vues de 20 personnes. Les images sont en niveaux de gris et de résolution 220×220. Elles sont tirées de séquences durant lesquelles les sujets tournent lentement la tête du profil à une vue frontale (voir figure A.5). Les sujets représentées sont de différents sexes, âges et origines ethniques.

FIG. A.5 – *Extrait de la base UMIST. Ensemble des vues disponibles pour la personne étiquetée sous le nom 1a.*

A.6 La base AR

La base AR [MB98] a été constituée en 1998 au sein du laboratoire Computer Vision Center (CVC) à Barcelone, en Espagne. 116 personnes (63 hommes et 53 femmes) sont enregistrées. Les images sont en couleur, de taille 768×576 pixels. 26 vues de chacun de ces sujets ont été collectées lors de deux sessions, menées à deux semaines d'intervalle. Lors de chaque session, 13 vues par personne ont été enregistrées. Un extrait des images collectées lors de la première session est fourni en figure A.6. Ces vues présentent des changements d'expression faciale, d'illumination, ainsi que des occultations partielles des yeux (port de lunettes) et de la partie basse du visage (cache-nez). Lors de la seconde session, les 13 vues sont collectées dans les mêmes conditions que pour la première session.

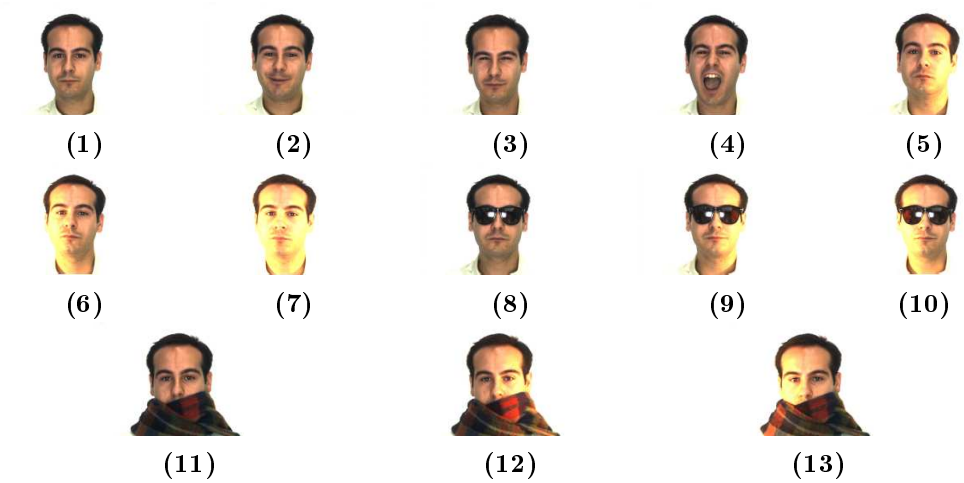


FIG. A.6 – *Extrait de la base AR. Ensemble des vues collectées lors de la première session, pour l'une des cent-seize personnes représentées.*

A.7 La base PIE

La base PIE [SBB03] a été collectée au sein de la Carnegie Mellon University entre octobre et décembre 2000. La base contient 41368 images de 68 personnes. Les images sont en couleur et de taille 640×480 pixels. Les angles de prise de vue, les conditions d'illumination ainsi que les expressions faciales sont systématiquement variées. Les vues ont été collectées dans une salle spécifique (*CMU 3D*) équipée de 13 appareils photographiques synchronisés de haute qualité et de 21 flashs. La figure A.7 montre les vues d'un sujet sous les 13 angles de prise de vue différents.

En plus des variations de pose, quatre autres facteurs ont été pris en compte :

- *Illumination I* (voir figure A.8-gauche) : les 21 flashs sont activés séquentiellement de manière très rapide. Les lumières de la pièce sont allumées ;
- *Illumination II* (voir figure A.8-droite) : les 21 flashs sont activés de la même manière que pour l'illumination I, mais avec les lumières de la pièce éteintes ;
- expression : les expressions représentées sont : expression neutre, sourire, clignement d'un œil et les deux yeux fermés. Les images collectées par les 13 caméras sont disponibles dans la base ;
- parole : 60 vues de chaque personne en train de parler sont enregistrées sous trois angles de prise de vue différents (de face, de trois-quarts et de profil).

FIG. A.7 – *Extrait de la base PIE. Variations de pose du profil droit (c22) au profil droit (c34), en passant par la pose frontale (c27).*

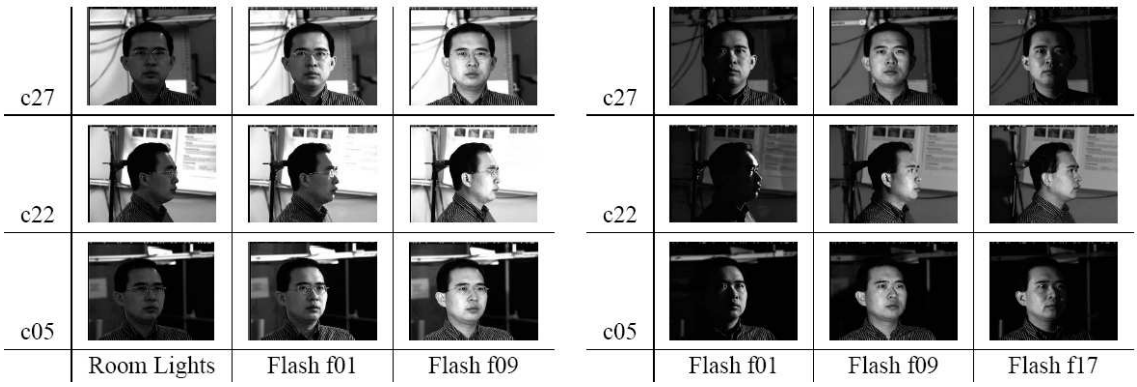


FIG. A.8 – *Extrait de la base PIE : à gauche des vues Illumination I et à droite des vues Illumination II. À gauche on visualise les effets des différents flashs avec les lumières allumées et à droite avec les lumières éteintes.*

Annexe B

La détection de visages et de caractéristiques faciales

B.1 Détection de visages

Il existe deux bases de référence couramment utilisées pour l'évaluation de la plupart des algorithmes de détection des visages. Il s'agit de la base CMU [RBK98], composée de 130 images contenant en tout 507 visages, et de la base du MIT [SP98]. Des sous-bases excluant les dessins de visages sont également couramment utilisées pour l'évaluation : il s'agit des bases CMU-125 et MIT-20.

Les premiers travaux portant sur la détection de visages dans des images à fond simple ou complexe remontent au milieu des années 1990. Des avancées très importantes ont été réalisées ces dernières années. Deux états de l'art détaillés et relativement récents sont fournis dans [HL01, YKA02]. Tout comme les techniques de reconnaissance, les approches de détection peuvent être divisées en deux catégories : les approches *locales* et les approches *globales*.

Les approches locales les plus anciennes reposent sur une analyse bas niveau de l'image par l'étude des coins, de l'intensité, de la couleur ou du mouvement. D'autres approches utilisent la mise en correspondance de modèles locaux, statiques ou déformables. Ces modèles sont généralement localisés autour des caractéristiques faciales (yeux, nez, etc.) et nécessitent donc la détection de celles-ci. Une fois les caractéristiques détectées, elles sont organisées de manière à définir un modèle plus global des visages, en tenant compte d'un ensemble de contraintes géométriques. Ces techniques reposent sur le choix d'un bon compromis entre influence de l'information globale et des modèles locaux, de manière par exemple à fournir au système une certaine robustesse aux occultations partielles des visages. La recherche des caractéristiques faciales étant menées dans l'intégralité de l'image, elle est fortement influencée par le bruit de l'image. Par conséquent, l'imprécision et le nombre de faux positifs (en termes de détection de caractéristiques faciales) sont potentiellement importants, ce qui nuit aux performances de la détection de visages.

Les méthodes globales, c.-à-d. basées sur la détection du visage dans sa globalité, ont été introduites dans le but d'être appliquées à des images contenant plusieurs visages et/ou en présence d'un fond complexe. Ces techniques évitent les problèmes d'imprécision dans la détection des caractéristiques faciales par le biais de l'apprentissage des règles intrinsèques des visages. Pour cela, il est nécessaire de disposer d'un volume de données important et présentant une variabilité importante (p. ex. dans les conditions de prise de vue). Ces dernières approches sont en général plus robustes au bruit et à des déformations de la région faciale que les approches locales. Pour la plu-

part, elles traitent la détection comme un problème de classification binaire, où les deux classes sont les visages et les non-visages. Les approches globales les plus anciennes reposent sur l'utilisation de techniques d'analyse statistique multivariée [CH97, SP98, SK00, GT00, YKA01, Liu03] pour la représentation des régions de l'image et/ou leur classification. Cette phase d'analyse est éventuellement précédée d'une phase de *clustering* (*cf.* note de bas de page n° 8 en p. 40) non supervisé des données d'entrée (*algorithme des K-moyennes* [JD88] ou *cartes de Kohonen* [Koh89]), et/ou suivies de classifieurs Bayésiens ou basés sur des réseaux de neurones. D'autres approches reposent sur l'utilisation directe de réseaux de neurones ou de Machines à Vecteurs de Support (voir section 2.2.5) pour la classification [OFG97, RBK98, FBVC01, VJ01, GD04].

Parmi les trois techniques reportées comme étant les plus performantes (*cf.* tableau B.1), on compte la technique de Féraud *et al.* [FBVC01], celle de Viola et Jones [VJ01] et celle de Garcia et Delakis [GD04].

Détecteur de visages	**CMU**	**CMU-125**	**MIT**	**MIT-20**
Colmenarez et Huang [CH97]	93,9%/8122			
Sung et Poggio [SP98]			79,9%/5	
Schneiderman et Kanade [SK00]		94,4%/65		
Yang *et al.* [YKA01]		93,6%/74		91,5%/1
Rowley *et al.* [RBK98]	86,2%/23		84,5%/8	
Féraud *et al.* [FBVC01]	86%/8			
Viola et Jones [VJ01]	88,4%/31		77,8%/5	
Garcia et Delakis [GD04]	90,3%/8	90,5%/8	90,1%/7	90,2%/5

TAB. B.1 – *Extrait de [GD04]. Évaluation des performances des principales méthodes de reconnaissance de visages (Taux de détection/Nombre de faux positifs), sur les bases CMU et MIT.*

La méthode de Féraud *et al.* est basée sur l'utilisation d'un type particulier de réseaux de neurones qualifié de *génératif contraint* et entraîné sur les valeurs de pixels des images globales. Le réseau est un Perceptron Multi-Couches auto-associatif et entièrement connecté. Il est conçu de manière à mettre en œuvre une Analyse en Composantes Principales (ACP) non-linéaire (voir section 2.2.2.6). Il est qualifié de « génératif », en ce sens qu'il fournit une estimation de la probabilité que le modèle ait généré le signal d'entrée. Le qualificatif « contraint » lui est appliqué, car des contre-exemples sont utilisés lors de l'apprentissage pour améliorer la qualité du modèle. Plusieurs réseaux génératifs contraints sont construits et combinés suivant un modèle de mélange conditionnel, de manière à pouvoir détecter des visages sous des poses non frontales et à réduire le nombre de faux positifs. Afin de réduire le coût de calcul, des préfiltrages basés sur la détection de teinte chair et la segmentation de mouvement sont mis en œuvre. Comme le montre la table B.1, cette approche a la particularité de fournir des bons taux de détection, avec

un nombre de faux positifs très faible.

La technique proposée par Viola et Jones repose sur une cascade de classifieurs rééchantillonnés, construits depuis des caractéristiques locales. Des caractéristiques robustes à des changements de luminance et de contraste sont extraites à différentes positions et à différentes échelles dans l'image originale. Ces caractéristiques sont inspirées des fonctions de base des ondelettes de Haar, puisqu'elles dérivent de la différence entre sommes de pixels de régions rectangulaires adjacentes de l'image. Un modèle d'*intégration* des images globales permet leur extraction rapide. Pour plus de détails, se référer à [VJ01]. L'ensemble des caractéristiques ainsi obtenues étant redondant, les plus importantes sont sélectionnées par le biais de l'algorithme de *boosting* [FS96] (*cf.* annexe G, p. 205) appelé Adaboost [FS97], qui permet de former à partir des caractéristiques sélectionnées un classifieur linéaire. Dans le but de rendre la détection plus rapide, une cascade de classifieurs de complexité croissante est construite. Chaque classifieur peut rejeter définitivement une région candidate. Seules les régions qui ne sont rejetées par aucun des classifieurs dans la cascade sont considérées comme des visages. La cascade est organisée de manière à ce que les éléments de fond soient rapidement rejetés, et les régions ressemblant le plus à des visages examinées plus en profondeur. On construit un modèle par pose de la tête. En phase de détection, on applique un classifieur de pose (*cf.* note de bas de page en p. 15) en amont de la détection. Un détecteur spécialisé dans la pose frontale est capable de traiter 15 images de taille 720×576 par seconde sur un processeur P4 de 3,2GHz. Le processus de détection de pose, suivi de la détection du visage, pourrait traiter environ 7 images par seconde. Les taux de détection garantis dans cette méthode sont excellents, mais sur certaines bases le taux de faux positifs peut être relativement élevé (*cf.* table B.1).

La technique de Garcia et Delakis, appelée *Convolutional Face Finder* (CFF), est basée sur l'utilisation de réseaux de neurones *convolutionnels* multi-couches. Le réseau, entraîné de manière supervisée depuis une base d'apprentissage contenant des images de visages et de non-visages, est capable de dériver automatiquement des extracteurs de caractéristiques (produits de convolution) spécialisés. Le réseau, illustré en figure B.1, comporte six couches, les quatre premières servant à l'extraction des caractéristiques et les deux dernières à la classification. Les signaux d'entrée du réseau sont des fenêtres de taille 32×36 pixels extraites de chacune de ces images redimensionnées. Dans la première couche (C_1), les fenêtres sont convoluées par des noyaux de taille 5×5 avec l'ajout d'un biais ; quatre noyaux différents sont appliqués, chacun résultant en une carte de caractéristiques faciales différente. Puis, dans une deuxième couche (S_1), les cartes ainsi obtenues sont sous-échantillonnées d'un facteur deux, pondérées et corrigées par l'ajout d'un biais. Cette opération permet de réduire la dimensionnalité du problème et d'améliorer la robustesse aux translations, rotations, changements d'échelle et déformations des images de visages. Les troisième et quatrième couches C_2 et S_2 consistent à réitérer les mêmes opérations, mais avec des noyaux de taille 3×3. Les deux dernières couches contiennent de simples perceptrons sigmoïdes et visent à effectuer la classification (visage/non-visage). Les neurones de la couche N_1 sont entièrement connectés à ceux de la couche S_2. La couche N_2, elle, ne contient qu'un seul neurone, connecté à tous ceux de la couche N_1. Le réseau est entraîné à l'aide d'un algorithme de rétropropagation adapté aux réseaux de neurones convolutionnels, décrit dans [LCBD+90], de manière à ce que la sortie de N_2 soit égale à 1 pour des visages et à -1 pour des non-visages. Cette technique globale présente plusieurs avantages par rapport aux précédentes. Premièrement, elle ne repose pas sur une phase d'extraction de caractéristiques coûteuse et potentiellement imprécise. De plus, il est inutile de procéder à une détection de pose en amont. Elle diffère également des méthodes précédentes par l'unicité du classifieur. Celui-ci est capable de détecter des visages tournés de $\pm 20°$ dans le plan et de $\pm 60°$ en profondeur. Le CFF est caractérisé par le meilleur ratio taux de détection / nombre de faux positifs reporté à ce

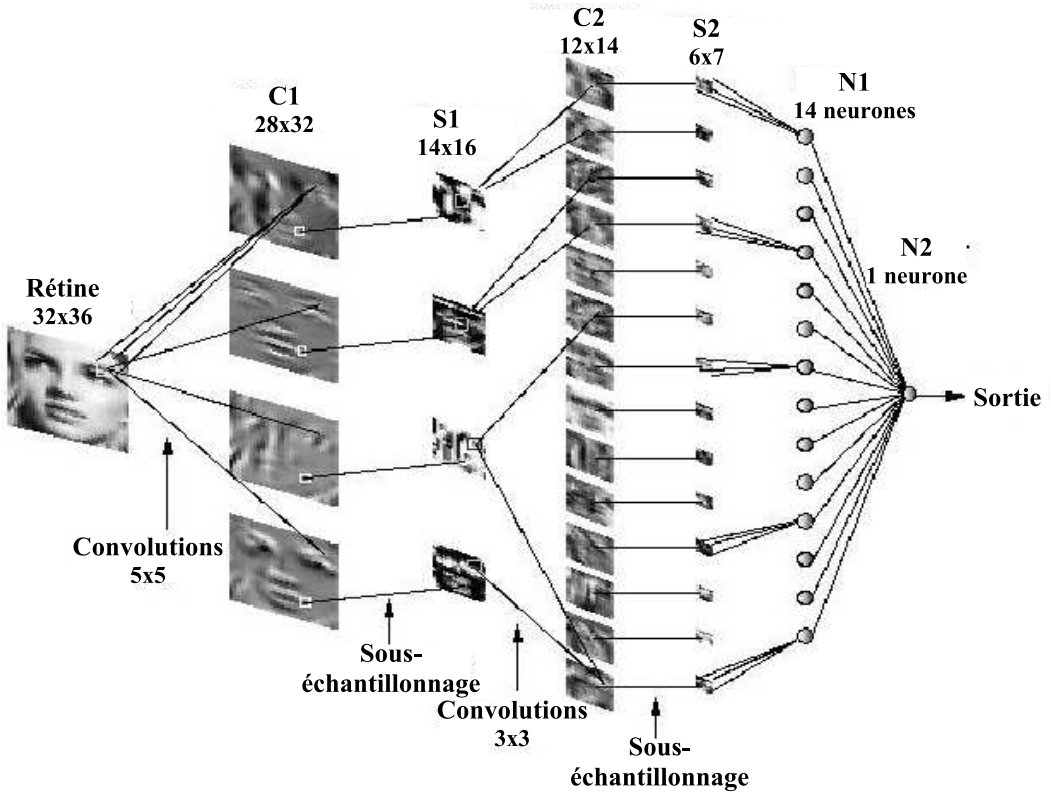


FIG. B.1 – *Adapté de [GD04]. Architecture du réseau de neurones convolutionnel à six couches (CFF) proposé par Garcia et Delakis.*

jour (*cf.* table B.1). La rapidité d'exécution est d'environ 4 images de taille 384×288 par seconde, avec un processeur P4 de 1,6GHz. Le CFF permet d'obtenir d'excellents taux de détection et un nombre de faux positifs très faible (*cf.* table B.1).

B.2 Détection de caractéristiques faciales

La plupart des approches de détection de caractéristiques faciales utilisent de manière indépendante un ensemble de détecteurs, chacun étant spécialisé pour une caractéristique donnée. Le mode de mise en œuvre de ces approches rejoint celui des méthodes locales de détection de visages (*cf.* section B.1).

Les Modèles Actifs d'Apparence (MAA), détaillés en section 2.2.3, ont plus récemment été introduits [CC04]. Ils visent à prédire les localisations des caractéristiques faciales qui permettent de faire correspondre au mieux la région faciale et un modèle de visage préalablement appris. Ce modèle est basé sur la forme et la texture des visages. Les MAA présentent l'avantage d'utiliser des contraintes géométriques (modèle de forme) durant la détection, et non en aval de celle-ci. Mais cette approche repose sur une procédure d'optimisation coûteuse et instable.

Très récemment, Duffner et Garcia ont proposé une méthode globale très efficace, basée sur l'utilisation de réseaux de neurones convolutionnels (*cf.* section B.1) [DG05]. Cette technique permet de détecter rapidement et précisément les caractéristiques faciales de manière robuste jusqu'à $\pm 60°$ dans le plan de l'image et $\pm 30°$ en profondeur, dans des images à fond complexe.

Annexe C

Normalisation des visages dans les images

Cette annexe vise à décrire le processus de normalisation appliqué sur les images de visages en amont de la reconnaissance. La normalisation consiste à aligner tous les visages de la même manière dans leur image associée. Elle est nécessaire pour garantir de bonnes performances dans la plupart des systèmes de reconnaissance (*cf.* section 1.6.3). On dispose des positions des caractéristiques faciales (yeux, nez et bouche) dans l'image. On considère que les images de visages sont en niveaux de gris. Si cela n'est pas le cas, on transforme les images de couleur en images à 256 niveaux de gris (en général, il suffit de moyenner les valeurs des trois canaux RGB).

Si les visages sont représentés sous une pose frontale (on tolère environ 15° de rotation en profondeur), on peut utiliser une méthodologie proche de celle proposée dans le cadre du protocole FERET [PMRR00]. Cette technique compte quatre étapes, illustrées en figure C.1 :

1. *rotation* du visage dans l'image, de manière à ce que l'axe interoculaire soit horizontal ;
2. *changement de l'échelle* de l'image, de manière à ce que le segment interoculaire soit de 70 pixels. Si la taille initiale de cet axe est supérieure à 70 pixels, on effectue un sous-échantillonnage du visage de facteur $\frac{I_i}{70}$, où I_i est la distance interoculaire initiale. Si, par contre, la distance interoculaire dans l'image initiale est inférieure à 70 pixels, on met en œuvre une interpolation bicubique [BJL03] ;
3. *découpage* de l'image à une résolution finale de 150×130 pixels, de manière à ce que le visage soit centré dans l'image. Pour cela, les coordonnées de l'œil droit (à gauche dans l'image) dans la base (x,y) d'origine le coin en haut à gauche de l'image est $(30, 45)$ (*cf.* figure C.1) ;
4. *égalisation de l'histogramme* de chacune des images ainsi obtenues. L'égalisation d'histogramme a pour but d'harmoniser la répartition des niveaux de luminosité de l'image, de manière à tendre vers un même nombre de pixels pour chacun des niveaux de gris de l'histogramme [BJL03]. Cette opération vise à gommer les différences entre images dues à des changements dans les conditions d'illumination.

Dans le contexte de certaines applications, une résolution inférieure à 150×130 pixels sera utilisée (soit par nécessité car les images sont trop petites, soit pour des raisons de rapidité de traitement). Dans ce cas, on diminue la distance interoculaire à l'étape 2. et le reste du processus est mis en œuvre de manière à conserver le même ratio entre distance interoculaire et résolution de l'image. La position de l'oeil droit dans l'image obtenue est redéfinie en fonction de cette nouvelle distance interoculaire.

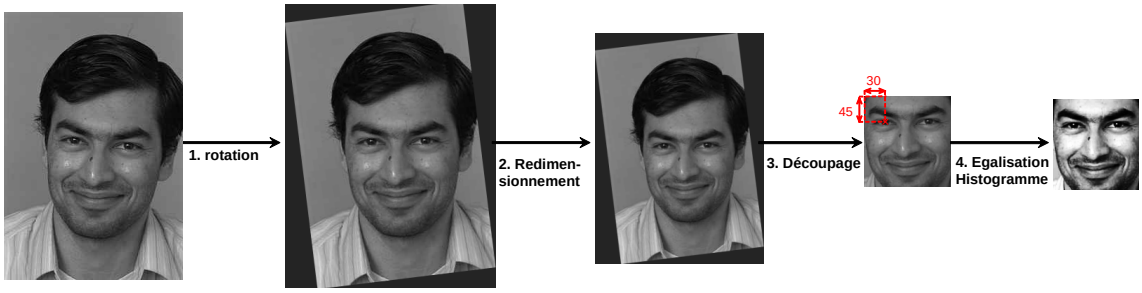


FIG. C.1 – *Le processus de normalisation des visages compte quatre étapes. Il nécessite la connaissance des positions des caractéristiques faciales dans l'image et est un préalable nécessaire à la majorité des systèmes de reconnaissance.*

Notons que, si la pose en profondeur est supérieure à $\pm 15°$, le changement d'échelle et le centrage du visage dans l'image ne doivent plus se faire uniquement à l'aide des positions des yeux, car sinon les images normalisées pourraient contenir des visages à des résolutions très différentes. Dans ce cas, il faut également tenir compte des positions des autres caractéristiques faciales, notamment du nez et/ou de la bouche.

Annexe D

Les mesures de dissimilarité usuelles

D.1 Les distances de Minkowski et leurs extensions fractionnaires

Notons $X = [X_1, X_2, \ldots, X_n]^T$ et $Y = [Y_1, Y_2, \ldots, Y_n]^T$ deux vecteurs de $\mathbb{R}^n$. Les distances d_{L_p} de Minkowski entre les vecteurs X et Y sont définies par la formule générale :

$$d_{L_p}(X,Y) = \sqrt[p]{\sum_{i=1}^{n} |X_i - Y_i|^p} \tag{D.1}$$

où $p \in \mathbb{N}$.

Pour $p = 1$, la distance de Minkowski d_{L_1} est appelée *distance de Manhattan* :

$$d_{L_1}(X,Y) = \sum_{i=1}^{n} |X_i - Y_i| \tag{D.2}$$

Pour $p = 2$, la distance de Minkowski d_{L_2} est appelée *distance Euclidienne* :

$$d_{L_2}(X,Y) = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2} \tag{D.3}$$

Si $p \in]0\,,1[$, la mesure de dissimilarité d_{L_p} n'est plus une distance car l'inégalité triangulaire n'est pas vérifiée. Cependant, l'utilisation de ces mesures a été justifiée expérimentalement. Ces mesures sont notamment réputées très efficaces pour la classification de données de grandes dimensions [AHK01, LDGF04]. Afin d'illustrer leurs propriétés, la figure D.1 montre les boules unités associées.

D.2 La mesure d'angle

La mesure de dissimilarité d'angle négatif (ou du cosinus) entre les vecteurs X et Y est définie comme suit :

$$d_{\cos}(X,Y) = 1 - \frac{X^T Y}{\|X\|\|Y\|} = 1 - \frac{\sum_{i=1}^{n} X_i Y_i}{\sqrt{\sum_{i=1}^{n} X_i^2 \sum_{i=1}^{n} Y_i^2}}. \tag{D.4}$$

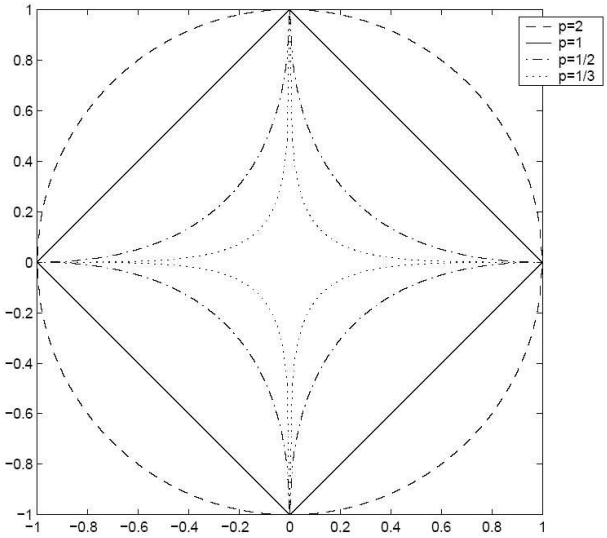


FIG. D.1 – *Boules unité en deux dimensions associées aux mesures de dissimilarité L_p (pour p allant de $1/3$ à 2)*

D.3 La mesure de divergence de Kullback-Leibler

La *mesure de divergence de Kullback-Leibler* (qui n'est pas une distance puisqu'elle est non symétrique et ne satisfait pas l'inégalité triangulaire) est une mesure de dissimilarité entre deux densités de probabilités f_1 et f_2 :

$$\delta(f_1,f_2) \quad = \quad \int f_1(y) \log\left(\frac{f_1(y)}{f_2(y)}\right) dy \tag{D.5}$$

Notons que cette mesure peut s'étendre à des distributions discrètes en remplaçant les intégrales par des sommes (sous la condition qu'aucune probabilité ne soit nulle).

D.4 Les mesures de dissimilarité utilisées pour les *eigenfaces*

Les mesures de dissimilarité suivantes sont généralement utilisées pour la classification des signatures issues de la technique des *eigenfaces* [Yam00]. Notons λ_i les valeurs propres de la matrice S_T, rangées dans leur ordre décroissant, et X' et Y' les signatures (projections) associées aux vecteurs X et Y :

La *distance de Mahalanobis-L_1* d_{Mah_1} est définie comme suit :

$$d_{\text{Mah}_1}(X',Y') \quad = \quad \sum_{i=1}^{n} \frac{1}{\sqrt{\lambda_i}} |X'_i - Y'_i| \tag{D.6}$$

La *distance de Mahalanobis-L_2* d_{Mah_2} s'écrit :

$$d_{\text{Mah}_2}(X',Y') \quad = \quad \sqrt{\sum_{i=1}^{n} \frac{1}{\lambda_i} (X'_i - Y'_i)^2} \tag{D.7}$$

La *distance de Mahalanobis-cosinus* $d_{\cos}$ s'écrit :

$$d_{\cos}(X', Y') = 1 - \frac{\sum_{i=1}^{n} \frac{1}{\lambda_i} X'_i Y'_i}{\sqrt{\sum_{i=1}^{n} \frac{1}{\lambda_i} {X'_i}^2 \sum_{i=1}^{n} \frac{1}{\lambda_i} {Y'_i}^2}}. \tag{D.8}$$

La distance proposée par Moon dans [MP98] et souvent appelée *distance de Moon*, est la suivante :

$$d_{\text{Moon}}(X', Y') = \sum_{i=1}^{n} \sqrt{\frac{\lambda_i}{\lambda_i + \alpha^2}} X'_i Y'_i \tag{D.9}$$

où α est une constante.

La *distance de Yambor*, introduite dans [YDB00], s'écrit :

$$d_{\text{Yamb}}(X', Y') = \sum_{i=1}^{n} \frac{1}{\sqrt{\lambda_i}} X'_i Y'_i \tag{D.10}$$

D.5 Les mesures de dissimilarité utilisées pour les *fisherfaces*

Dans [Zha99], Zhao a introduit une mesure de dissimilarité spécialement conçue pour la comparaison de signatures X' et Y' issues de la technique des *fisherfaces*, appelée *distance des fisherfaces souple* :

$$d^{(\alpha)}(X', Y') = \sum_{i=1}^{n} \lambda_i^{\alpha} (X'_i - Y'_i)^2 \tag{D.11}$$

où la valeur du paramètre α est strictement positive, et λ_i est la valeur propre associée à la $i^{\text{ème}}$ *fisherface*.

Annexe E

Description de l'ADL

La présente annexe est organisée comme suit. La section E.1 donne une formulation géométrique de l'ADL ; un point de vue probabiliste est adopté en section E.2. La section E.3 détaille les principaux algorithmes de mise en œuvre de l'ADL. Les notations utilisées sont celles du chapitre 3 (table 3.1, p. 57).

E.1 Formulation géométrique : l'Analyse Factorielle Discriminante

On recherche de nouvelles variables, appelées *axes discriminants* et en nombre g, correspondant à des directions de $\mathbb{R}^n$ qui séparent le mieux possible en projection les k groupes d'observations. Il existe au plus $k-1$ axes discriminants permettant de séparer linéairement les k classes (voir figure 3.2). Supposons $\mathbb{R}^n$ muni d'une métrique M. Nous reviendrons plus tard sur le choix de la métrique M. Notons w_i les axes discriminants de $\mathbb{R}^n$. On appelle *facteur discriminant* associé à l'axe w_i le vecteur W_i de $\mathbb{R}^n$ tel que $W_i = Mw_i$. Il s'agit de la forme linéaire, donnant la coordonnée de la projection M-orthogonale sur l'axe w_i. Notons $w = [w_1, w_2, \ldots, w_g]$ la matrice de $\mathbb{R}^{n\times g}$ contenant en colonne les axes discriminants, rangés par pouvoir discriminant décroissant, et $W = [W_1, W_2, \ldots, W_g] \in \mathbb{R}^{n\times g}$ la matrice des facteurs discriminants associés. Supposons de plus que la matrice w est M-orthonormée : $w^T M w = I_g$, où I_g est la matrice identité de taille $g \times g$. La projection X' de $X \in \mathbb{R}^n$ sur w est donc donnée par :

$$X' =< w, X >_M = W^T X \tag{E.1}$$

On appellera *variables discriminantes* les projections $W^T A$ de la matrice des observations A sur les axes discriminants.

E.1.1 Critère à optimiser

On considérera qu'un facteur W_i est discriminant s'il minimise par projection les variations à l'intérieur des classes, tout en maximisant les variations entre classes. Le pouvoir discriminant d'un facteur W_i est donc mesuré par le critère ci-dessous (plus $J(W_i)$ est grand, plus W_i est discriminant) :

$$J(W_i) \quad = \quad \frac{S_b'^{(i)}}{S_w'^{(i)}} \tag{E.2}$$

où les matrices $S_b'^{(i)}$ et $S_w'^{(i)}$ sont respectivement les matrices de variance inter- et intra- classes des données projetées sur l'axe W_i :

$$S_b'^{(i)} = W_i^T S_b W_i \qquad \text{et} \qquad S_w'^{(i)} = W_i^T S_w W_i \tag{E.3}$$

La mesure (E.2) du pouvoir discriminant du facteur W_i devient donc :

$$J(W_i) \quad = \quad \frac{W_i^T S_b W_i}{W_i^T S_w W_i} \tag{E.4}$$

Dans le cas de deux classes, on recherche le vecteur W_1 de $\mathbb{R}^n$ maximisant le critère (E.4). Sous l'hypothèse que S_w est régulière, on peut considérer sans perte de généralité que cela revient à maximiser $W_1^T S_b W_1$ sous la contrainte $W_1^T S_w W_1 = 1$. On est donc en présence d'un problème de maximisation sous contrainte ; l'annulation de la dérivée du lagrangien par rapport à W_1 conduit au problème d'analyse propre généralisé suivant :

$$S_b W_1 \quad = \quad \lambda_1 S_w W_1 \tag{E.5}$$

où λ_1 est un scalaire tel que $0 \leq \lambda_1 < +\infty$. Si la matrice S_w est inversible, alors le facteur discriminant est le vecteur propre de $S_w^{-1} S_b$ associé à la plus grande valeur propre λ_1. Il n'est cependant pas nécessaire de résoudre le système propre, car $S_b W_1$ est toujours dans la direction de la droite $(\overline{A_1} - \overline{A_2})$ [15]. À un facteur de normalisation près et sous l'hypothèse que la matrice de variation intra-classe S_w est inversible, on obtient :

$$W_1 \quad = \quad S_w^{-1}(\overline{A_1} - \overline{A_2}) \tag{E.6}$$

C'est la *fonction de Fisher*, introduite par Fisher en 1936 [Fis36].

Ce n'est qu'en 1948 que Rao a étendu la définition de l'Analyse Discriminante Linéaire au cas de $k > 2$ classes [Rao48]. On recherche le sous-espace $\mathcal{F}$ de $\mathbb{R}^n$, linéaire et de dimension $g \leq k-1$, dans lequel les classes sont le mieux séparées possible. Ce sous-espace sera appelé dans la suite *sous-espace discriminant*. Il nous faut donc généraliser la mesure de séparabilité (E.4) à un critère qui porte non pas sur le pouvoir discriminant d'un facteur, mais d'un sous-espace $\mathcal{F}$ entier. On peut montrer que les matrices de variance intra-classe S_w' et inter-classe S_b' des données projetées sur $\mathcal{F}$ peuvent s'écrire :

$$S_w' = W^T S_w W \qquad \text{et} \qquad S_b' = W^T S_b W \tag{E.7}$$

et on a $S_T' = S_w' + S_b'$. Deux types de mesures numériques reflétant la dispersion, à savoir la trace et le déterminant de la matrice de variance, ont servi à l'élaboration de critères dans le cas multi-classes. On peut notamment citer les critères suivants (à maximiser) :

$$J_1(W) \quad = \quad \text{trace}(S_T'^{-1} S_b') \tag{E.8}$$

$$J_2(W) \quad = \quad \text{trace}(S_w'^{-1} S_b') \tag{E.9}$$

$$J_3(W) \quad = \quad \frac{|S_b'|}{|S_T'|} \tag{E.10}$$

$$J_4(W) \quad = \quad \frac{|S_b'|}{|S_w'|} \tag{E.11}$$

15. Résultat facile à montrer si l'on remarque que, dans le cas à deux classes, la matrice de variance inter-classe S_b peut se réécrire $S_b = \frac{N_1 N_2}{N}(\overline{A_1} - \overline{A_2})(\overline{A_1} - \overline{A_2})^T$.

où $|X|$ est le déterminant de la matrice X. Sous l'hypothèse que la matrice S_w est inversible, il n'existe pas réellement d'indicateur permettant d'orienter son choix vers un critère plutôt qu'un autre. En revanche, le critère le plus couramment utilisé dans le domaine de la reconnaissance de formes est le critère (E.11) [DHS01], souvent appelé *critère de Fisher*. De la même manière que dans le cas à deux classes, on peut montrer que maximiser le critère (E.11) revient à maximiser le déterminant de la matrice $W^T S_b W$, sous la contrainte $W^T S_w W = I_g$ [Fuk90]. Sous l'hypothèse que la matrice S_w est inversible, les colonnes de la matrice W maximisant le critère (E.11) sont les vecteurs propres de la matrice $S_w^{-1} S_b$ associés aux plus grandes valeurs propres (non nulles). Étant donné que la matrice S_b est la somme de k matrices $(\overline{A_j} - \overline{A})(\overline{A_j} - \overline{A})^T$, chacune de ces matrices étant de rang inférieur ou égal à un, et qu'au plus $k-1$ de ces matrices sont linéairement indépendantes (les données sont liées par la relation $\frac{1}{N}\sum_{j=1}^{k} N_j \overline{A_j} = \overline{A}$), le rang de la matrice S_b est au plus $k-1$. Par conséquent, il existe au plus $k-1$ vecteurs propres correspondant à des valeurs propres non nulles. On choisit $g \leq k-1$ vecteurs propres W_i associés aux plus grandes valeurs propres, pour former l'ensemble des facteurs discriminants que l'on stocke dans la matrice $W = [W_1, W_2, \ldots, W_g]$. Traditionnellement, le nombre g de facteurs discriminants considérés augmente avec le nombre N d'observations dont on dispose [DHS01]. Par souci de simplicité, nous considérerons dans la suite de cette section que les observations sont linéairement indépendantes, et que donc il existe exactement $k-1$ vecteurs propres correspondant à des valeurs propres non nulles. Étant donné que la matrice $S_w^{-1} S_b$ n'est pas nécessairement symétrique, le calcul de son système propre est potentiellement instable. Heureusement, il n'est pas forcément nécessaire de résoudre le système propre de $S_w^{-1} S_b$ pour obtenir les facteurs discriminants. Plus de détails sont donnés en section E.3.

E.1.1.1 Choix de la métrique

Revenons au choix de la métrique M. Les facteurs discriminants (et donc les variables discriminantes) sont indépendants du choix de la métrique. La métrique $M = S_w^{-1}$, dite de Mahalanobis, vérifiant la condition de M-orthogonalité de l'axe discriminant, elle est souvent retenue. On peut facilement montrer que maximiser le critère (E.4) revient à maximiser le critère :

$$J(W) \quad = \quad \frac{W^T S_b W}{W^T S_T W} \tag{E.12}$$

où la matrice de variance intra-classe a été remplacée au dénominateur par la variance totale S_T, donnée en équation (3.6). Sous l'hypothèse que la matrice S_T est inversible, les facteurs discriminants W_i sont les vecteurs propres de $S_T^{-1} S_b$ associés aux plus grandes valeurs propres μ_i. On peut facilement montrer que les vecteurs propres de $S_T^{-1} S_b$ sont les mêmes que ceux de $S_w^{-1} S_b$, et que leurs valeurs propres associées μ_i vérifient la relation :

$$\mu_i = \frac{\lambda_i}{1 + \lambda_i}$$

Il est donc équivalent de maximiser le critère (E.4) ou le critère (E.12) ; par conséquent, le choix de la métrique $M = S_w^{-1}$ ou de la métrique $M = S_T^{-1}$ est indifférent.

E.1.1.2 Signification des valeurs propres

Intéressons-nous à l'interprétation de la valeur propre μ_i, comprise entre 0 et 1 :

– $\mu_i = 1$ correspond au cas suivant :

$$\left(S_T^{-1} S_b W_i = W_i\right) \Rightarrow \left(S_T S_T^{-1} S_b W_i = S_T W_i\right) \Rightarrow (S_b W_i = (S_w + S_b) W_i) \Rightarrow (S_w W_i = 0_{n \times 1})$$

où $0_{n\times 1}$ est le vecteur de $\mathbb{R}^n$ ne contenant que des zéros. Par conséquent, la matrice $W_i^T S_w W_i$, de dispersion intra-classe des données M-projetées sur w_i, est nulle. Chacun des k nuages est donc dans un hyperplan M-orthogonal à w_i. Les classes sont parfaitement séparées si les centroïdes se projettent sur w_i en des points différents (voir figure E.1) ;

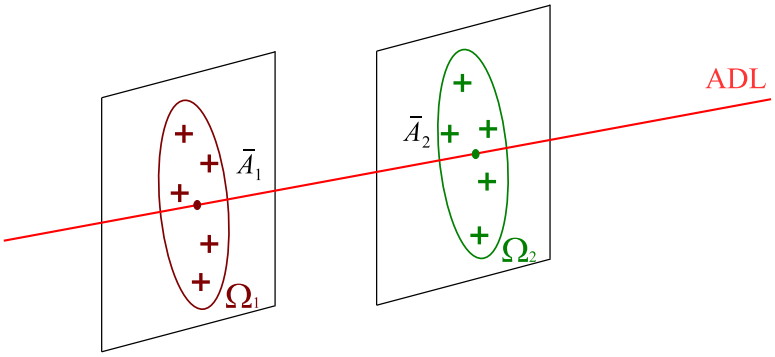


FIG. E.1 – *Cas où $\mu_i = 1$. Les dispersions des groupes projetés sur l'axe discriminant sont nulles. Chacun des k nuages repose donc dans un hyperplan orthogonal à l'axe discriminant. Les classes sont parfaitement séparées si les centroïdes des classes projetées sont distincts.*

– *a contrario*, si $\mu_i = 0$, on a :

$$\left(S_T^{-1} S_b W_i = 0_{n\times 1}\right) \Rightarrow \left(S_T S_T^{-1} S_b W_i = 0_{n\times 1}\right) \Rightarrow \left(S_b W_i = 0_{n\times 1}\right)$$

Par conséquent, la matrice de variance inter-classe $W_i^T S_b W_i$ des données projetées sur w_i est nulle. Les centres de gravité des classes sont confondus : c'est par exemple le cas où les nuages sont concentriques (voir figure E.2). Aucune séparation linéaire n'est possible, mais il peut exister une fonction non-linéaire permettant de séparer les deux nuages : dans le cas illustré en figure E.2 on peut citer par exemple la distance au centre, qui est une fonction quadratique des variables.

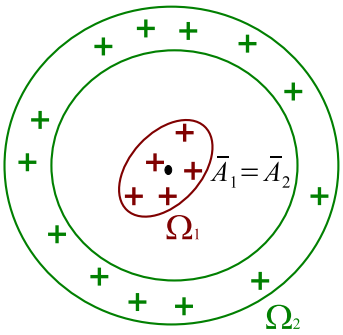


FIG. E.2 – *Cas où $\mu_i = 0$. Les moyennes des deux groupes sont confondues. Aucune règle linéaire ne permet de séparer les deux groupes.*

La figure E.3 illustre le fait qu'un facteur associé à une valeur μ_i inférieure à 1 peut néanmoins séparer parfaitement les deux nuages. La valeur propre μ_i est donc une mesure pessimiste du

pouvoir discriminant du facteur associé W_i.

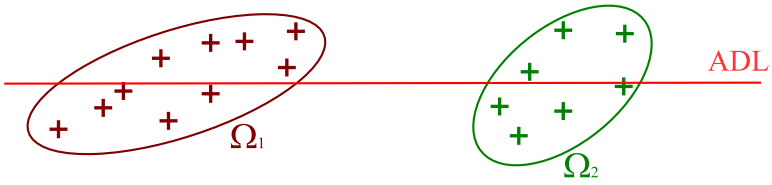


FIG. E.3 – *Cas où $\mu_i \neq 1$ (les matrices de dispersion intra-classe des groupes projetés ne sont pas nulles). Pourtant, les deux groupes sont parfaitement séparés par l'axe discriminant.*

E.1.2 Classification des données

Une fois la matrice des facteurs discriminants W calculée, on peut considérer que le *classifieur* (modèle) est construit. Ce modèle, entièrement déterminé par sa matrice de facteurs W et les observations de la base d'apprentissage (étiquetées par leur classe), va nous permettre de prédire la classe d'appartenance de nouvelles observations. La classification des données se fait par le biais de calculs de distances entre éléments projetés dans le sous-espace discriminant $\mathcal{F}$. La méthodologie est la suivante. On dispose d'une nouvelle observation X, à laquelle on souhaite assigner une identité. On commence par calculer les coordonnées $X' = W^T X$ de X dans $\mathcal{F}$. Puis, on calcule la distance Euclidienne (au carré) $d^2_{\mathcal{F}}(X,\overline{A_j})$ entre X' et chacun des centres projetés $\overline{A_j}' = W^T \overline{A_j}$:

$$d^2_{\mathcal{F}}(X,\overline{A_j}) \quad = \quad (X' - \overline{A_j}')^T (X' - \overline{A_j}'), \qquad \text{pour } \; j = 1, 2, \ldots, k \tag{E.13}$$

On décide d'affecter X à la classe Ω^*_j de distance minimum :

$$\Omega^*_j \quad = \quad \underset{j=1,\ldots,k}{\text{Argmin}} \left[d^2_{\mathcal{F}}(X,\overline{A_j}) \right] \tag{E.14}$$

En développant la distance $d^2_{\mathcal{F}}(X,\overline{A_j})$ on trouve :

$$d^2_{\mathcal{F}}(X,\overline{A_j}) = X'^T X' + \overline{A_j}'^T \overline{A_j}' - 2X'^T \overline{A_j}'$$

Étant donné que $X'^T X'$ ne dépend pas de la classe Ω_j, la règle d'affectation peut donc se réécrire :

$$\Omega^*_j \quad = \quad \underset{j=1,\ldots,k}{\text{Argmin}} \left[\overline{A_j}'^T \overline{A_j}' - 2X'^T \overline{A_j}' \right] \tag{E.15}$$

On voit que cette règle est linéaire par rapport aux coordonnées de X dans $\mathcal{F}$. Pour chaque individu à classer, il faut donc : 1) le projeter dans $\mathcal{F}$, ce qui demande un nombre d'opérations en $o(ng)$, et 2) calculer k fonctions linéaires de ses g coordonnées (soit un nombre d'opérations en $o(kg)$) pour en chercher la valeur minimale, soit un nombre total d'opérations en $o((n+k)g)$.

Sous l'hypothèse que la matrice S_w est inversible, on peut montrer que la distance Euclidienne entre deux éléments X et Y de $\mathbb{R}^n$, projetés dans $\mathcal{F}$, est équivalente à la distance de Mahalanobis dans l'espace initial $\mathbb{R}^n$ suivante :

$$d^2_{\mathcal{F}}(X,Y) \quad = \quad (X-Y)^T S_w^{-1} (X-Y) \tag{E.16}$$

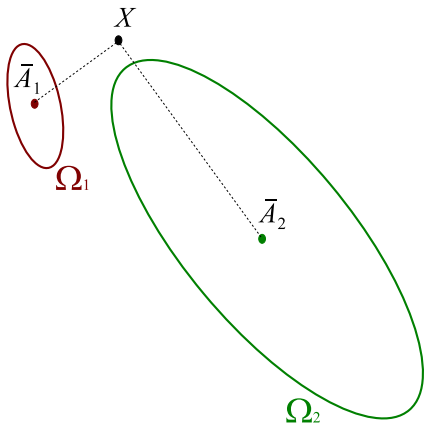


FIG. E.4 – *Bien que X soit plus proche de $\overline{A_1}$ que de $\overline{A_2}$, l'observation X appartient probablement à la deuxième classe, du fait de la forme des nuages de points.*

Le critère à minimiser (E.15) peut donc se réécrire :

$$\Omega_j^* = \underset{j=1,\ldots,k}{\text{Argmin}} \left[\overline{A_j}^T S_w^{-1} \overline{A_j} - 2X^T S_w^{-1} \overline{A_j} \right] \tag{E.17}$$

Cette règle est linéaire par rapport aux coordonnées de X. Pour chaque individu à classer, il faut donc calculer k fonctions linéaires de ses n coordonnées et en chercher la valeur minimale (ce qui demande un nombre d'opérations en $o(kn)$).

Supposons que $g = k - 1$. Si $n \gg k^2 - k$, alors $(n + k)g \ll kn$. Étant donné que, pour la plupart des applications de reconnaissance de formes, et *a fortiori* pour la reconnaissance de visages, on a $n \gg k^2 - k$, on préférera généralement, pour des raisons de coût de calcul, utiliser la formulation (E.15) plutôt que le critère (E.17) basé sur la distance de Mahalanobis.

E.1.3 Insuffisance des règles géométriques

L'utilisation de la règle précédente conduit à des affectations incorrectes lorsque les dispersions des groupes sont très différentes entre elles : rien ne justifie alors l'usage de la même métrique pour toutes les différentes classes (voir figure E.4). Cela suggère que le classifieur construit par Analyse Discriminante Linéaire n'est optimal que sous un certain nombre de conditions, que nous détaillerons dans la suite de cette section.

E.2 Formulation probabiliste

E.2.1 La règle Bayésienne

Notons $f_j(X)$ la densité de probabilité associée à l'observation $X \in \mathbb{R}^n$, issue de la classe Ω_j. Connaissant l'observation X, la probabilité *a posteriori* que X provienne du groupe Ω_j est donnée par la formule de Bayes :

$$\mathbb{P}[\Omega_j / X] = \frac{\mathbb{P}[\Omega_j] f_j(X)}{\sum_{i=1}^{k} \mathbb{P}[\Omega_i] f_i(X)} \tag{E.18}$$

où les $\mathbb{P}[\Omega_i]$ sont les probabilités *a priori* d'appartenance à chacune des classes (issues des proportions de chaque classe dans la population totale). La règle Bayésienne consiste à assigner à

X l'identité de la classe ayant la plus forte probabilité *a posteriori*. Les dénominateurs étant identiques pour chacune des k classes, l'observation A sera affectée à la classe Ω_j^* telle que

$$\Omega_j^* = \underset{j=1,\ldots,k}{\text{Argmax}} \left(\mathbb{P}[\Omega_j] f_j(X)\right) \tag{E.19}$$

Cette règle, appelée *règle Bayésienne*, minimise la probabilité d'erreur dans le contexte d'un problème de classification avec des distributions des classes $f_j(X)$ connues ou estimées. Il existe un certain nombre de méthodes dites *non paramétriques* [Sap90], au sens où elles ne nécessitent pas d'hypothèse spécifique sur la famille de loi de probabilité. Nous nous focaliserons ici sur des techniques *paramétriques*, où l'on considère que les densités f_j sont issues d'une famille de lois de probabilités, dont les observations de la base d'apprentissage vont nous servir à estimer les paramètres.

E.2.2 Le modèle multinormal

On suppose que les observations de chaque groupe Ω_j sont distribuées selon une loi $\mathcal{N}(\mu_j,\Sigma_j)$; on obtient :

$$f_j(X) = \frac{1}{(2\pi)^{n/2}|\Sigma_j|^{1/2}} \exp\left[-\frac{1}{2}(X-\mu_j)^T\Sigma_j^{-1}(X-\mu_j)\right] \tag{E.20}$$

où $|\Sigma_j|$ est le déterminant de la matrice de variance-covariance Σ_j. Généralement, Σ_j est estimée par son estimateur asymptotiquement sans biais $\frac{N_j}{N_j-1}V_j$ et μ_j par $\overline{A_j}$. En passant en logarithme, la règle (E.19) revient à chercher la classe Ω_j^* telle que :

$$\Omega_j^* = \underset{j=1,\ldots,k}{\text{Argmin}} \left[(X-\mu_j)^T\Sigma_j^{-1}(X-\mu_j) - 2\ln(\mathbb{P}[\Omega_j]) + \ln(|\Sigma_j|)\right]$$

Lorsque les Σ_j sont différentes d'une classe à l'autre cette règle est quadratique et donne naissance à la technique dite d'*Analyse Discriminante Quadratique* (ADQ). Dans ce cas, la règle de décision nécessite la comparaison de k fonctions quadratiques de X.

E.2.3 Le modèle multinormal homoscédastique

On parle d'*homoscédasticité* si les matrices de variance-covariance des différents groupes sont égales deux à deux : $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_k = \Sigma$. Dans ce cas, $\ln(|\Sigma_j|)$ est constante et $(X-\mu_j)^T\Sigma^{-1}(X-\mu_j)$ est égale à la distance de Mahalanobis entre X et μ_j. Si l'on développe la règle de Bayes (E.19) sans prendre en compte $X^T\Sigma^{-1}X$ ni $\ln(|\Sigma|)$, qui ne dépendent pas de la classe, on obtient :

$$\Omega_j^* = \underset{j=1,\ldots,k}{\text{Argmax}} \left[X^T\Sigma^{-1}\mu_j - \frac{1}{2}\mu_j^T\Sigma^{-1}\mu_j + \ln(\mathbb{P}[\Omega_j])\right] \tag{E.21}$$

où l'espérance μ_j et la matrice Σ sont respectivement estimées par $\overline{A_j}$, et par l'estimateur asymptotiquement sans biais $\frac{N}{N-k}S_w$. On remarque qu'alors, sous la condition que chaque classe d'appartenance ait la même probabilité d'être observée, la règle Bayésienne (E.21) est strictement équivalente à la règle géométrique (E.15).

On peut donc en conclure que, sous les conditions que chaque groupe d'observations soit distribué suivant une loi multinormale et que les matrices de variance-covariance des différents groupes soient égales, l'Analyse Discriminante Linéaire permet de construire un classifieur optimal au sens de la règle de Bayes.

E.3 Algorithmes de construction du modèle

Dans cette section, nous présentons les deux algorithmes principaux permettant de calculer la matrice W des facteurs discriminants.

E.3.1 La procédure de résolution standard

Nous avons vu que, sous l'hypothèse que la matrice de variance intra-classe S_w est régulière, les facteurs discriminants peuvent être calculés par une analyse propre de la matrice $S_w^{-1}S_b$. La matrice W est telle que $W = [W_1, \ldots, W_g]$, où les $g \leq k-1$ facteurs discriminants W_i sont les vecteurs propres de $S_w^{-1}S_b$, rangés par ordre décroissant de leur valeur propre associée. Étant donné que la matrice $S_w^{-1}S_b$ n'est pas nécessairement symétrique, le calcul de son système propre est potentiellement instable. Une technique basée sur les diagonalisations des matrices S_w et S_b, symétriques réelles, est détaillée ci-après.

E.3.2 La procédure de résolution par diagonalisations (algorithme de Fukunaga)

Fukunaga [Fuk90] a montré que la matrice W optimale au sens du critère de Fisher (E.11), de taille $n \times (k-1)$, vérifie les conditions suivantes :

$$W^T S_w W = I_{k-1} \quad \text{et} \quad W^T S_b W = \Lambda \tag{E.22}$$

où I_{k-1} est la matrice identité de taille $(k-1) \times (k-1)$, et Λ est une matrice diagonale de déterminant maximal. On dit que la matrice W diagonalise simultanément les matrices S_w et S_b. W peut être calculée par une technique en deux étapes, basée sur les diagonalisations successives de la matrice S_w et d'une matrice transformée de S_b, selon l'algorithme ci-après [Fuk90, SW96].

1. on détermine les éléments propres (V, D_V) de la matrice de dispersion intra-classe S_w. La matrice orthonormée V, de taille $n \times n$, est constituée en colonnes des vecteurs propres de S_w, et D_V est la matrice diagonale contenant les valeurs propres associées.
   Puis, on projette les centres des classes $\overline{A_j}$ sur $VD_V^{-1/2}$, selon $\overline{A_j}' = D_V^{-1/2}V^T\overline{A_j}$;
2. on applique une ACP sur ces centres projetés $\overline{A_j}'$. Notons $\overline{A}' = \frac{1}{N}\sum_{j=1}^{k} N_j\overline{A_j}'$ la moyenne des k centres projetés. On peut aisément montrer que la matrice de dispersion S_b' des centres projetés, telle que :

   $$S_b' = \frac{1}{N}\sum_{j=1}^{k} N_j(\overline{A_j}' - \overline{A}')(\overline{A_j}' - \overline{A}')^T$$

   peut se réécrire :

   $$S_b' = D_V^{-1/2}V^T S_b V D_V^{-1/2} \tag{E.23}$$

   Puisque S_b est symétrique réelle, S_b' l'est aussi. Donc, S_b' est diagonalisable et il existe une base de vecteurs propres orthonormée. Le rang de la matrice S_b vérifie : $\text{rang}(S_b) \leq \min(n, k-1)$. Pour la plupart des applications, on a $n > k-1$. Par conséquent, le rang de S_b' est au plus $k-1$. Il existe donc au plus $k-1$ vecteurs propres de S_b' associés à des valeurs propres non nulles. Ceux-ci sont stockés dans la matrice B', orthonormée et de taille $n \times (k-1)$;

3. Les facteurs discriminants de la matrice W globale de l'ADL sont définis comme suit :

$$W = VD_V^{-1/2}B' \tag{E.24}$$

Vérifions que la matrice W ainsi construite satisfait bien les conditions de diagonalisations simultanées données en équation (E.22) :

$$\begin{array}{rclcrclcl} W^T S_w W & = & B'^T D_V^{-1/2} V^T S_w V D_V^{-1/2} B' & = & B'^T I_n B' & = & B'^T B' & = & I_{k-1} \\ W^T S_b W & = & B'^T D_V^{-1/2} V^T S_b V D_V^{-1/2} B' & = & B'^T S'_b B' & = & \Lambda & & \end{array}$$

où Λ est la matrice de taille $(k-1) \times (k-1)$, telle que $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_{k-1})$, où les λ_i sont les valeurs propres non nulles de la matrice S'_b, rangées par ordre décroissant. On peut également montrer que la matrice W obtenue par diagonalisations successives est la même que celle obtenue par la méthode standard (c'est-à-dire par analyse propre de la matrice $S_w^{-1}S_b$). Notons que l'algorithme de Fukunaga, tout comme l'algorithme de résolution standard, nécessite que la matrice S_w soit de rang plein (car sinon la matrice D_V contient des valeurs propres nulles et n'est pas inversible à l'étape 1.).

Lien avec l'ACP Notons que les vecteurs-images, projetés sur $VD_V^{-1/2}$, ont une matrice de dispersion intra-classe unitaire. En effet, leur matrice de covariance intra-classe S'_w peut s'écrire :

$$\begin{array}{rcl} S'_w & = & D_V^{-1/2} V^T S_w V D_V^{-1/2} \\ \Rightarrow S'_w & = & I_n \end{array} \tag{E.25}$$

L'égalité (E.25) provient de la définition du système propre orthonormé (V, D_V) de la matrice S_w. On dit parfois que l'étape 1. du processus ci-dessus est une phase de « blanchiment » (ou de normalisation) des données au sens de leur variabilité intra-classe. On peut également noter que l'étape 2. consiste à appliquer une ACP sur les centres « blanchis » (c'est-à-dire projetés sur la matrice $VD_V^{-1/2}$).

Annexe F

L'ADL sous-optimale

F.1 L'ADL dans le noyau

Chen *et al.* [CLK$^+$00] ont montré que le noyau de S_w peut contenir de l'information discriminante, si la projection de S_b est non nulle dans les directions ainsi définies. Notons W une telle direction. Par définition, on a

$$W^T S_w W = 0 \quad \text{et} \quad W^T S_b W \geq 0 \tag{F.1}$$

ce qui implique que le critère de Fisher donné en équation (3.11) atteint un maximum global dans cette direction :

$$J(W) = \frac{|W^T S_b W|}{|W^T S_w W|} = \infty$$

Les principales techniques permettant de construire une ADL dans le noyau de S_w et introduites dans la littérature sont détaillées dans cette section.

F.1.1 L'ADL$_0$

La technique d'ADL dans le noyau proposée par Chen [CLK$^+$00] et que nous désignerons par la suite par le sigle ADL$_0$, peut être résumée de la manière suivante :

1. diagonalisation de la matrice S_w. Puisque S_w est symétrique réelle, on peut choisir ses vecteurs propres de manière à ce qu'ils forment une base V orthonormée. La matrice singulière S_w est de rang $r \leq N - k < n$. Considérons que les vecteurs propres V_i de $V = [V_1, \ldots, V_r, \ldots V_n]$ sont rangés dans V par ordre décroissant de leur valeur propre associée. Regroupons les vecteurs propres associés à des valeurs propres nulles dans la matrice $Q = [V_{r+1}, \ldots V_n]$. Notons $M = n - r$. La matrice Q, de taille $n \times M$, forme une base du noyau de S_w et donc on a :

   $$Q^T S_w Q = 0_M$$

   où 0_M est la matrice carrée de taille $M \times M$ ne contenant que des '0'.
   Puis, les centres des classes $\overline{A_j}$ sont projetés dans le noyau de S_w, pour obtenir les vecteurs $\overline{Z_j}$ de $\mathbb{R}^M$, tels que :

   $$\overline{Z_j} = Q^T \overline{A_j} \quad \forall\, j = 1, \ldots, k \tag{F.2}$$

2. application d'une ACP sur les centres projetés dans le noyau de S_w et pondérés par les effectifs de leurs classes. La matrice de variance S'_b de ces centres projetés peut s'écrire comme suit :

$$S'_b = \frac{1}{N}\sum_{j=1}^{k} N_j(\overline{Z_j} - \overline{Z})(\overline{Z_j} - \overline{Z})^T = Q^T S_b Q \tag{F.3}$$

où $\overline{Z} = \frac{1}{N}\sum_{j=1}^{k} N_j \overline{Z_j}$ est la moyenne pondérée de ces centres.
On sélectionne les $g = k - 1$ vecteurs propres B'_i de S'_b associés à des valeurs propres non nulles ; on les stocke dans la matrice orthonormée B'. Les valeurs propres associées sont rangées dans la matrice diagonale $D_{B'}$, par ordre décroissant ;
3. la matrice de projection W de l'ADL globale est alors définie comme suit :

$$W = QB' \tag{F.4}$$

On a donc obtenu la matrice W telle que :

$$W^T S_w W = B'^T Q^T S_w Q B' = B'^T 0_M B' = 0_g \tag{F.5}$$
$$W^T S_b W = B'^T Q^T S_b Q B' = B'^T S'_b B' = D_{B'} \tag{F.6}$$

qui correspond à une valeur du critère de Fisher (3.11) très grande.

Cette technique nécessite le calcul du rang de la matrice S_w, ce qui est une opération mal posée. Le deuxième désavantage majeur de cette technique est que, quand la taille de S_w est très grande – ce qui est généralement le cas pour la reconnaissance de visages –, le calcul des vecteurs propres de S_w est très coûteux et instable numériquement. Notamment, le calcul des vecteurs propres de S_w associés aux valeurs propres nulles ou proches de zéro est généralement imprécis. C'est pourquoi Chen *et al.* [CLK+00], procèdent en amont de leur technique à une agglomération de pixels, permettant d'extraire des caractéristiques géométriques. Or, rien ne dit que cette phase préliminaire n'engendre pas de perte d'information discriminante. De plus, des expérimentations [CNWB05] montrent que, plus la taille du noyau de S_w est importante, plus la technique de Chen est performante. Par conséquent, toute phase de prétraitement conduisant à une réduction des dimensions de l'espace original (et donc du noyau de S_w) devrait être évitée.

F.1.2 L'ACP+ADL$_0$

Afin de réduire la complexité du calcul d'une base Q du noyau de S_w, Huang *et al.* ont proposé dans [HLLM02] une nouvelle technique, basée sur l'idée qu'il n'est pas nécessaire de conserver toute l'information provenant du noyau de S_w, car seule une partie de cette information est discriminante. Cette technique est désignée par le sigle ACP+ADL$_0$.

Notons $\mathrm{Ker}(X)$ le noyau de l'espace engendré par la matrice X. Huang *et al.* remarquent que $\mathrm{Ker}(S_T) = \mathrm{Ker}(S_w) \cap \mathrm{Ker}(S_b)$. En effet, posons Q un élément quelconque de $\mathrm{Ker}(S_T)$. Puisque $S_T = S_w + S_b$, on a :

$$\begin{aligned} Q^T S_T Q = 0 \quad &= \quad Q^T S_w Q + Q^T S_b Q \\ \Rightarrow Q^T S_w Q = 0 \qquad &\text{et} \qquad Q^T S_b Q = 0 \end{aligned} \tag{F.7}$$

car les matrices S_w et S_b sont toutes deux semi-définies positives. Donc, Q appartient à la fois à $\mathrm{Ker}(S_w)$ et à $\mathrm{Ker}(S_b)$. Les éléments Q de $\mathrm{Ker}(S_T)$ annulent donc simultanément le numérateur et le dénominateur du critère de Fisher (3.11). Ils ne participent donc pas à la maximisation de

ce critère. Par conséquent, l'espace $\mathrm{Ker}(S_T)$ ne contient pas d'information discriminante au sens du critère de Fisher.

Huang *et al.* proposent donc de neutraliser $\mathrm{Ker}(S_T)$ par une phase d'ACP (en retenant tous les vecteurs propres associés à des valeurs propres non nulles), en amont de la technique d'ADL_0 présentée en section F.1.1. Cette phase préliminaire d'ACP engendre un coût calculatoire supplémentaire. Les résultats expérimentaux de [HLLM02], obtenus sur la base ORL (*cf.* annexe A) montrent une amélioration des taux de reconnaissance par rapport à l'ADL_0, quand le nombre d'images par classe est faible ($N_j \leq 7$), c.-à-d. quand les dimensions du noyau de S_w sont importantes.

F.1.3 La méthode des *Vecteurs Discriminants Communs*

Cevikalp *et al.* ont proposé dans [CNWB05] une méthode baptisée méthode des *Vecteurs Discriminants Communs* et notée VDC, qui est très efficace en termes de coût de calcul et contourne le problème de l'évaluation des vecteurs propres de S_w associés aux plus petites valeurs propres. Cette technique consiste à extraire un vecteur (appelé Vecteur Discriminant Commun (VDC)) représentatif de chaque classe, en rejetant toutes les directions correspondant à des valeurs propres non nulles de la matrice de variance de la classe. L'hypothèse de l'homoscédasticité des classes est supposée vérifiée et donc seul le noyau de S_w doit être évalué. L'algorithme est le suivant :

1. calculer les vecteurs propres de S_w (de grande taille $n \times n$), associés à des valeurs propres non nulles (puisque $n \gg N$, on peut utiliser l'astuce rappelée en section 3.3.3.2 et se ramener à l'analyse propre d'une matrice de taille $N \times N$). On obtient ainsi la matrice $V = [V_1, \ldots, V_r]$ des vecteurs propres, où r est le rang de la matrice S_w. La matrice de projection dans le noyau de S_w peut être définie comme suit :

$$Q = I_n - VV^T \tag{F.8}$$

   Celle-ci permet de mesurer l'erreur de reconstruction, au sens de la norme Euclidienne ;
2. pour chaque classe $(\Omega_j)_{j=1\ldots k}$, choisir au hasard l'un de ses représentants $A_l \in \Omega_j$ et le projeter dans le noyau de S_w, de manière à obtenir le vecteur discriminant c_j associé à Ω_j :

$$\forall j = 1, \ldots, k, \quad \forall A_l \in \Omega_j, \quad c_j = A_l - VV^T A_l \tag{F.9}$$

   On peut montrer que, quelle que soit l'observation A_l de Ω_j choisie, le vecteur discriminant c_j est le même ;
3. notons $S_c = Y_c Y_c^T$ la matrice de dispersion associée à ces vecteurs discriminants communs, avec $Y_c = [c_1 - \bar{c}, c_2 - \bar{c}, \ldots, c_k - \bar{c}]$ et $\bar{c} = \frac{1}{k}\sum_{j=1}^{k} c_j$ est la moyenne des vecteurs communs. Puis, on calcule les vecteurs propres de S_c correspondant à des valeurs propres non nulles, en utilisant la même astuce qu'à l'étape 1. Il existe au plus $k-1$ de ces vecteurs W_i ; ils forment la matrice $W = [W_1, \ldots, W_{k-1}]$ de projection globale de l'ADL ainsi construite.

Ainsi, le critère de Fisher (3.11) est transformé selon :

$$W = \underset{|W^T S_w W = 0|}{\mathrm{Argmax}} |W^T S_b W| \tag{F.10}$$

$$W = \underset{|W^T S_w W = 0|}{\mathrm{Argmax}} |W^T S_T W| \tag{F.11}$$

$$W = \underset{W}{\mathrm{Argmax}} |W^T S_c W| \tag{F.12}$$

Le critère (F.10), déjà présenté dans [BHK97, BLP02], assure la maximisation de la dispersion entre classes dans le noyau de S_w. Dans [CNWB05], Cevikalp *et al.* présentent également un algorithme basé sur l'utilisation de sous-espaces et l'orthogonalisation de Gram-Schmidt, permettant de déterminer les *Vecteurs Communs Discriminants* sans avoir à manipuler de grosses matrices telles que S_w, ce qui diminue le temps de calcul et peut améliorer la stabilité de l'algorithme.

F.1.4 Synthèse

La technique d'ADL$_0$ présentée en section F.1.1 nécessite de calculer directement une base du noyau de S_w. Parce que les grandes dimensions du problème rendent cette opération difficile, une phase préalable d'extraction de caractéristiques est appliquée. Cette étape préliminaire est non souhaitable, car elle pourrait conduire à de la perte d'information discriminante. C'est pourquoi Huang *et al.* [HLLM02] proposent de la remplacer par une phase d'ACP retenant tous les axes principaux de valeur propre non nulle (technique ACP+ADL$_0$, présentée en section F.1.2). Si une telle ACP ne conduit pas à la perte d'information discriminante (au sens du critère de Fisher), elle est néanmoins coûteuse. Enfin, Cevikalp *et al.* [CNWB05] proposent un algorithme de résolution (la méthode des VDC présentée en section F.1.3) moins coûteux que l'ADL$_0$ et l'ACP+ADL$_0$.

Ces trois approches utilisent l'information contenue dans le noyau de S_w. Par contre, elles ne prennent en compte aucune information en dehors du noyau de S_w. Néanmoins, des résultats expérimentaux (notamment ceux de Yu et Yang détaillés dans la section ci-après) suggèrent qu'il existe de l'information discriminante (en termes de capacité de généralisation de la méthode) en dehors du noyau de S_w.

F.2 L'ADL Directe

Les techniques d'ADL dans le noyau de S_w présentées en section F.1 ne prennent pas en compte l'information – potentiellement discriminante – située en dehors de celui-ci. De plus, ces méthodes nécessitent de déterminer et de manipuler le noyau de S_w, difficilement évaluable. Afin de pallier ces inconvénients, Yu et Yang ont introduit dans [YY01] une technique appelée ADL Directe (ADLD) et consistant à inverser l'ordre des diagonalisations par rapport à une ADL$_0$: on commence par blanchir les données selon S_b, avant de procéder à la diagonalisation de la matrice de covariance intra-classe S'_w dans l'espace des données blanchies. L'algorithme est le suivant :

1. on commence par calculer le système propre de S_b : notons B la matrice orthonormée des vecteurs propres de S_b associée aux g plus grandes valeurs propres, où $g \leq \text{rang}(S_b) = \min(n,k-1) = k-1$. Les valeurs propres sont stockées dans la matrice diagonale D_b, de taille $g \times g$. On obtient donc :

$$D_b^{-1/2} B^T S_b B D_b^{-1/2} = I_g \tag{F.13}$$

   Blanchissons les données au sens de leur matrice de variance inter-classe, selon :

$$Y_l = (B D_b^{-1/2})^T A_l = D_b^{-1/2} B^T A_l \quad \forall\, l = 1, \ldots, N \tag{F.14}$$

2. la matrice de covariance intra-classe S'_w des données blanchies Y_l est :

$$S'_w = \frac{1}{N} \sum_{j=1}^{k} \sum_{Y_l \in \Omega_j} (Y_l - \overline{Y_j})(Y_l - \overline{Y_j})^T = D_b^{-1/2} B^T S_w B D_b^{-1/2} \tag{F.15}$$

où $\overline{Y_j} = \frac{1}{N_j} \sum_{Y_l \in \Omega_j} Y_l$. On calcule son système propre, noté (V',$D_{V'}$), tel que les vecteurs propres de V' sont orthonormés ;

3. on peut optionnellement ne garder dans V' que les $g' < g$ vecteurs propres de S'_w associés à des valeurs propres nulles ;
4. les facteurs discriminants du classifieur d'ADL directe sont contenus dans la matrice W telle que :

$$W = BD_b^{-1/2}V' \tag{F.16}$$

Vérifions que la matrice W diagonalise simultanément le numérateur et le dénominateur du critère de Fisher (3.11) :

$$W^T S_w W = V'^T D_b^{-1/2} B^T S_w B D_b^{-1/2} V' = V'^T S'_w V' = D_{V'} \tag{F.17}$$

$$W^T S_b W = V'^T D_b^{-1/2} B^T S_b B D_b^{-1/2} V' = V'^T I_g V' = I_{g'} \tag{F.18}$$

Plus le déterminant de $D_{V'}$ est faible, plus le critère de Fisher (3.11) est important.

Les expérimentations de Yu et Yang [YY01] sont menées sur la base ORL (*cf.* Annexe A), aléatoirement divisée en une base d'apprentissage et une base de test, chacune contenant 5 images par personne. Cette opération est répétée dix fois et le taux de reconnaissance moyen est calculé. L'algorithme d'ADLD sans l'étape 3. donne un taux de reconnaissance de 90,8%, contre 86,6% avec l'étape 3. On voit donc que, si l'on rejette l'information hors de S_w, le taux de reconnaissance (86,6%) est bon. Néanmoins, il est significativement meilleur (90,8%) si l'on garde au moins une partie de cette information. Cela tend à prouver le noyau de S_w contient une grande partie de l'information discriminante, mais qu'il existe en dehors de cet espace de l'information utile pour la classification.

L'avantage principal de l'ADLD sur les techniques d'ADL dans le noyau (voir section F.1) est qu'elle tient compte de l'information discriminante en dehors du noyau de S_w. De plus, comme son nom l'indique, elle peut être appliquée directement sur les images de visages sans nécessiter d'étape préliminaire d'ACP ou d'agglomération de pixels.

Par contre, Yu et Yang font l'hypothèse que le noyau de S_b ne contient aucune information discriminante, ce qui est faux en général. En effet, cela revient à considérer que les vecteurs discriminants appartiennent au sous-espace engendré par les centres des classes. La figure F.1 montre un exemple où l'hypothèse de Yu et Yang n'est pas vérifiée.

F.3 L'ADL Duale

Les techniques duales visent à tenir compte à la fois de l'information provenant du noyau et de son complémentaire.

F.3.1 L'ADL Duale de Wang et Tang

Wang et Tang ont introduit dans [WT04a] une technique combinant deux ADL, l'une effectuée dans le sous-espace principal $\mathcal{W}$ de S_w, l'autre effectuée dans son complémentaire $\mathcal{W}^\perp$. Les deux ADL sont effectuées en parallèle et l'on dispose ainsi de deux modèles. Lors de la phase de classification, les deux modèles sont utilisés conjointement pour calculer une distance inspirée de celle introduite dans [MP97]. Cette distance est basée sur la combinaison de deux mesures

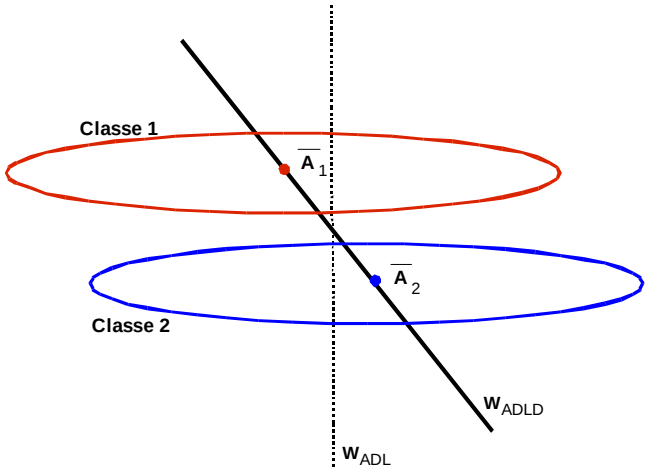


FIG. F.1 – *Exemple de deux classes multinormales de même matrice de covariance. Le vecteur discriminant de l'ADL Directe (noté W_{ADLD}) est contraint de passer par les deux centres et ne permet pas de séparer totalement les deux classes. Le vecteur discriminant selon le critère de Fisher, noté W_{ADL}, permet de séparer les deux classes.*

de dissimilarités : la *Distance From Feature Space* (DFFS) calculée dans $\mathcal{W}$, et la *Distance In Feature Space* (DIFS) estimée dans $\mathcal{W}^{\perp}$. L'algorithme proposé est le suivant :

1. calculer les vecteurs propres de S_w associés aux K plus grandes valeurs propres. Ces vecteurs propres V_i constituent la matrice $V = [V_1, \ldots, V_K]$. La matrice D_V des valeurs propres de S_w est la matrice diagonale contenant les valeurs propres, rangées en ordre décroissant : $D_V = \text{diag}(\lambda_1, \ldots, \lambda_K, \lambda_{K+1}, \ldots, \lambda_n)$. La matrice Λ est la matrice des valeurs propres associées à V : $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_K)$. Les valeurs propres $\{\lambda_{K+1}, \ldots, \lambda_n\}$ sont très faibles et très difficiles à estimer précisément. Elles sont supposées toutes égales et telles que $\lambda_{K+1} = \lambda_{K+2} = \cdots = \lambda_n = \rho$, où
$$\rho = \frac{1}{n-K} \sum_{i=K+1}^{n} \lambda_i^*$$
où les λ_i^* sont estimées par l'ajustement d'une fonction non-linéaire sur la portion du spectre des valeurs propres calculé précisément ;
2. les centres des classes sont projetés sur $\mathcal{W}$ et normalisés par les valeurs propres associées. On calcule la matrice de covariance inter-classe correspondante :
$$K_b^P = \Lambda^{-1/2} V^T S_b V \Lambda^{-1/2}$$
Une ACP est appliquée sur K_b^P : on retient ses l_p axes principaux, que l'on stocke dans la matrice Φ_P ;
3. on en déduit la matrice de projection W_P dans l'espace discriminant de $\mathcal{W}$:
$$W_P = V \Lambda^{-1/2} \Phi_P$$
4. Comme dans [CNWB05], on considère que la matrice de projection dans $\mathcal{W}^{\perp}$ est $I_n - VV^T$, où I_n est la matrice identité de taille I_n. Il est facile de montrer que la matrice de variance des centres projetés dans $\mathcal{W}^{\perp}$ peut s'écrire :
$$K_b^C = (I_n - VV^T)^T S_b (I_n - VV^T)$$

On note Φ_C la matrice des l_C vecteurs propres de K_b^C associés aux plus grandes valeurs propres ;

5. la matrice de projection dans l'espace discriminant de $\mathcal{W}^\perp$ est :

$$W_C = (I - VV^T)\Phi_C$$

Lorsqu'une image-requête X doit être reconnue, on calcule, pour chaque classe Ω_j de la base d'apprentissage, la mesure de dissimilarité suivante :

$$d(X,\Omega_j) = \|W_P^T X - W_P^T \overline{A_j}\| + \frac{1}{\rho}\|W_C^T X - W_C^T \overline{A_j}\| \tag{F.19}$$

où $\overline{A_j} = \frac{1}{N_j}\sum_{A_l \in \Omega_j} A_l$ est la moyenne des visages de la classe Ω_j. Le visage X se voit assigner l'identité Ω_j^* telle que $d(X,\Omega_j^*) = \min_{j=1,\ldots,k} d(X,\Omega_j)$.

Cette technique est théoriquement intéressante, puisqu'elle permet d'extraire conjointement de l'information discriminante de $\mathcal{W}$ et de $\mathcal{W}^\perp$. Par contre, le modèle nécessite en tout trois paramètres : K, l_P et l_C, et aucune heuristique de choix n'est fournie. De plus, le mode d'évaluation de ρ n'est pas clair, alors même que cette valeur ρ, très petite, a un impact important sur la distance (F.19) utilisée, donc sur les résultats de classification. Il faut également noter que, aux étapes 1. et 4., on doit procéder à la diagonalisation de matrices de très grande taille, ce qui est coûteux et instable numériquement. Les résultats expérimentaux semblent mettre en évidence sur une sous-base de FERET des performances légèrement meilleures que celles des techniques de Chen *et al.*, de Yu et Yang [YY01] et des *fisherfaces*. Néanmoins, le protocole expérimental retenu manque de clarté.

F.3.2 L'ADL Duale de Yang *et al.*

La technique de Wang et Tang est séduisante théoriquement, puisqu'à la différence des méthodes exposées en section F.1 et F.2 elle ne repose sur aucune hypothèse concernant la localisation de l'information discriminante. Mais elle est très difficile à mettre en œuvre et instable, ce qui peut nuire à ses performances. On peut de plus considérer qu'il existe de l'information en dehors du noyau de S_w, mais qu'il n'est pas forcément nécessaire de conserver toute l'information provenant de $\mathcal{W}^\perp$. Dans cette optique, Yang *et al.* ont introduit dans [YZY03] une nouvelle technique d'ADL Duale. Comme dans la méthode d'ACP+ADL$_0$ (*cf.* section F.1.2), une ACP est appliquée sur les images originales de manière à obtenir une matrice de projection $P = [P_1,\ldots,P_M]$ contenant en colonnes les $M < n$ vecteurs propres P_i de S_T associés aux valeurs propres non nulles $\{\lambda_1,\ldots,\lambda_M\}$ (où $M = \text{rang}(S_T)$). Puis, on diagonalise $S'_w = P^T S_w P$; on obtient ainsi une matrice orthonormée de vecteurs propres $V' = [V'_1, V'_2, \ldots, V'_n]$ rangées dans l'ordre décroissant de leurs valeurs propres. Notons $\mathcal{W} = \text{vect}(V'_1, \ldots, V'_r)$ l'image de S'_w, où $r = \text{rang}(S'_w)$ et $\mathcal{W}^\perp = \text{vect}(V'_{r+1}, \ldots, V'_n)$ son noyau. Tous les vecteurs de $\mathcal{W}$ maximisant la matrice de dispersion inter-classe sont retenus pour former la matrice de projection W de la technique globale. Si plus de vecteurs de projection sont nécessaires à une bonne classification, la matrice W est complétée avec des vecteurs propres provenant de $\mathcal{W}^\perp$. Cette technique permet de prendre en compte toute l'information provenant du noyau de S_w et, si c'est nécessaire, de compléter avec une partie de l'information en dehors du noyau. Cette technique est donc très intéressante. Mais la phase préliminaire d'ACP engendre un coût de calcul supplémentaire. De plus, la règle permettant de décider si de l'information de $\mathcal{W}^\perp$ est nécessaire n'est pas claire.

Annexe G

Les techniques de rééchantillonnage

G.1 Les techniques de rééchantillonnage usuelles et l'ADL

La technique de *boosting*, proposé par Freund et Shapire [FS96], est basée sur une repondération itérative des exemples issus de la base d'apprentissage lors de la construction du classifieur. Initialement, tous les exemples de la base d'apprentissage se voient attribuer le même poids. On construit un classifieur à partir de cette base d'apprentissage. Puis, on remet à jours les poids de la base d'apprentissage, en augmentant les poids des exemples mal classés par ce classifieur. Un second classifieur est construit à partir de cette nouvelle base pondérée, et les exemples mal classés par ce nouveau classifieur voient leurs poids augmenter, etc. On construit ainsi une cascade de classifieurs, chacun étant dépendant des résultats de classification de son prédécesseur. Puis, les résultats de classification de tous ces classifieurs sont agrégés, généralement par une règle de vote à la majorité, pondérée ou non. Le but principal de l'algorithme de *boosting* est d'arriver à classer correctement des exemples dits « difficiles », tels que ceux qui se trouvent initialement à la frontière entre deux classes. Par conséquent, le *boosting* est indiqué dans le cas où les observations situées aux frontière sont très porteuses d'information et reflètent les vraies distributions des classes, c'est-à-dire si la base d'apprentissage contient beaucoup d'exemples. Par contre, si l'on ne dispose pas de suffisamment d'exemples, deux cas de figure sont possibles : soit les données d'apprentissage sont parfaitement classées par le premier classifieur, et l'algorithme de *boosting* n'a pas lieu d'être, soit l'algorithme se focalise sur un petit nombre d'exemples, potentiellement non représentatifs des classes, et le classifieur « boosté » peut donner de moins bons résultats que le classifieur initial [SD02]. Le *boosting* n'est donc pas conseillé pour l'ADL dans le cadre de la reconnaissance de visages, où la plupart du temps trop peu d'exemples sont disponibles en regard de leur dimensionnalité.

Le *bagging*, proposé par Breiman [Bre96], consiste à agréger [KHDM98] les résultats de classification obtenus par plusieurs classifieurs, chacun de ces classifieurs étant construit sur un échantillon *bootstrap* [ET93] de la base d'apprentissage. Un échantillon *bootstrap* $\Omega^{(b)}$ est construit aléatoirement, avec remplacement, depuis un échantillon initial Ω ; il contient le même nombre d'observations que Ω ; ses observations proviennent de Ω mais certaines ont été répliquées plusieurs fois dans $\Omega^{(b)}$. Prenons l'exemple de l'ADL$^+$. Comme nous l'avons vu en section 3.3.3.4, la capacité de généralisation de la technique d'ADL est très dégradée si la taille N de la base d'apprentissage est proche de sa dimensionnalité n. Skurichina et Duin [SD02] ont montré qu'appliquer le *bagging* sur la technique de ADL$^+$ engendre un classifieur ayant des performances comparables à un simple classifieur ADL$^+$, mais entraîné avec moins d'exemples (voir figure G.1).

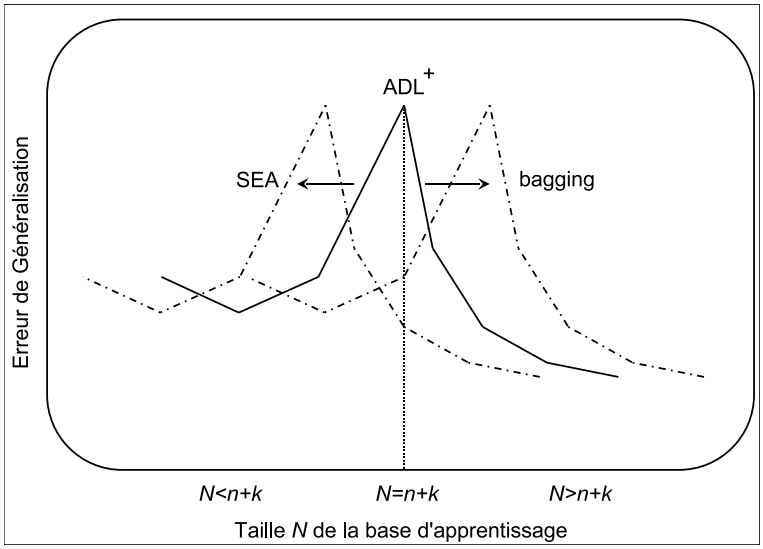


FIG. G.1 – *Adapté de [SD02]. Effets du bagging et de la méthode des Sous-Espaces Aléatoires sur la courbe d'apprentissage de l'ADL$^+$. Le classifeur « baggé » a un comportement proche d'un classifieur simple de même dimensionnalité, mais construit avec moins d'exemples (translation de la courbe d'apprentissage vers la droite), tandis que la méthode alliant SEA et ADL$^+$ se rapproche d'un classifieur ADL$^+$ simple, mais construit avec plus d'exemples (translation de la courbe d'apprentissage vers la gauche).*

Par conséquent, cette technique peut être efficacement utilisée pour des bases de taille critique (et en combinaison avec un classifieur surmontant le problème de la singularité, tel que l'ADL$^+$). Le *bagging* permet de définir un classifieur global moins sensible à la présence de valeurs aberrantes dans la base d'apprentissage. En effet, un certain nombre d'exemples de la base d'apprentissage ne sont pas inclus dans un échantillon *bootstrap*. En présence de valeurs aberrantes, utiliser le *bootstrap* augmente donc les chances de construire des classifieurs à partir de sous-bases moins bruitées que la base d'apprentissage originale, qui tireront les performances des plus bruités vers le haut lors de la phase d'agrégation [SD02].

La technique des Sous-Espaces Aléatoires (SEA) a été introduite par Ho [Ho98]. Il s'agit de construire plusieurs sous-échantillons de la base d'apprentissage initiale. Dans chaque sous-échantillon, tous les exemples de la base d'apprentissage sont sélectionnés, mais on ne considère qu'un nombre $p < n$ de leurs caractéristiques, sélectionnées aléatoirement parmi les n caractéristiques disponibles. Les résultats de classification de ces classifieurs sont ensuite combinés par une simple règle de vote à la majorité. Comme le montre Skurichina et Duin, appliquer la technique des sous-espaces aléatoires est utile pour l'ADL puisque le classifieur agrégé se comporte comme un classifieur simple, mais entraîné avec plus d'exemples (voir figure G.1). Par exemple, en combinaison avec l'ADL$^+$ (voir section 3.3.3.4), cela peut permettre d'améliorer les performances si la base d'apprentissage est de taille critique ou presque vide.

G.2 Le rééchantillonnage de l'ADL pour la reconnaissance de visages

G.2.1 La technique de Lu et Jain

Lu et Jain proposent dans [LJ03] un algorithme basé sur la technique des *fisherfaces* [BHK97] et un rééchantillonnage sans remplacement. Le rééchantillonnage est appliqué à l'intérieur de chaque classe de vecteurs-images Ω_j, de manière à ne garder dans le nouvel échantillon que $N'_j < N_j$ images, où N_j est le nombre d'images disponibles pour la classe Ω_j. Lu et Jain imposent que tous les N'_j soient égaux entre eux ($N'_1 = N'_2 = \cdots = N'_k$, où k est le nombre de classes de la base d'apprentissage). On crée ainsi B échantillons, contenant chacun $N^B = kN'_j$ vecteurs-images ; à partir de chaque échantillon $\Omega^{(b)}$, on construit un classifieur par la technique des *fisherfaces* (voir section 3.3.3.2). Les résultats de classification (obtenus par une mesure de cosinus au plus proche voisin) des B classifieurs sont agrégés à l'aide d'une règle de vote à la majorité. Le mode de fonctionnement de cette technique est schématisé en figure G.2.

Avec $B = 20$ sous-échantillons, les expérimentations montrent une amélioration sensible des taux de reconnaissance par rapport à un simple classifieur *fisherfaces* construit sur l'ensemble de la base d'apprentissage initiale. Il est à noter que la règle de la somme a également été évaluée pour l'agrégation et donne des résultats légèrement moins bons que le vote à la majorité. Ici, le rééchantillonnage, bien que sans remise, joue le même rôle que le *bagging*, à savoir qu'il permet de stabiliser le classifieur des *fisherfaces* par l'utilisation de plusieurs bases, dont certaines sont moins bruitées que la base initiale.

Cependant, chaque classifieur, construit depuis N^B vecteurs projetés de taille $N^B - k$, souffre d'instabilité. Les performances pourraient certainement être améliorées si moins d'*eigenfaces* étaient retenues en amont de l'ADL [SW96, LW98]. De plus, on peut également remarquer que cette technique est très coûteuse puisque, en tout, elle nécessite la mise en œuvre de B ACPs et B ADLs.

G.2.2 La méthode de Wang et Tang

Wang et Tang proposent dans [WT04b] deux techniques alliant variantes de l'ADL (surmontant le problème de la singularité) et méthodes de rééchantillonnage. Puis, ces deux techniques sont fusionnées pour donner naissance à une troisième méthode (voir figure G.3).

La première méthode introduite par Wang et Tang allie *fisherfaces* et SEA (voir figure G.3). Le classifieur issu de l'algorithme des *fisherfaces* de Belhumeur *et al.* [BHK97] est instable, car l'ADL est construite avec N visages projetés de dimension $N - k$; le nombre et la dimension des données sont proches, aussi l'ADL souffre-t-elle de surapprentissage. Pour améliorer les taux de reconnaissance, une solution serait de réduire la taille du sous-espace principal en retenant moins de vecteurs propres [SW96], mais cela peut engendrer une perte d'information discriminante. La solution de Wang et Tang [WT04b] consiste à appliquer une variante semi-aléatoire de la technique des sous-espaces aléatoires, entre les étapes d'ACP et d'ADL. La méthode peut se résumer ainsi :

1. appliquer une ACP sur les vecteurs-images. Retenir tous les $N-1$ vecteurs propres associés à des valeurs propres non nulles. Ces *eigenfaces* P_i sont stockées par ordre décroissant de leur valeur propre associée dans la matrice $P = [P_1, \ldots, P_{N-1}]$, orthonormée ;
2. générer B_1 sous-espaces $P^{(b)}$ à partir de la base P d'*eigenfaces*. Chaque sous-espace $P^{(b)}$ est constitué des N_0^B premières *eigenfaces* (de plus fortes valeurs propres), et de N_1^B *eigenfaces* choisies aléatoirement parmi les *eigenfaces* restantes ;

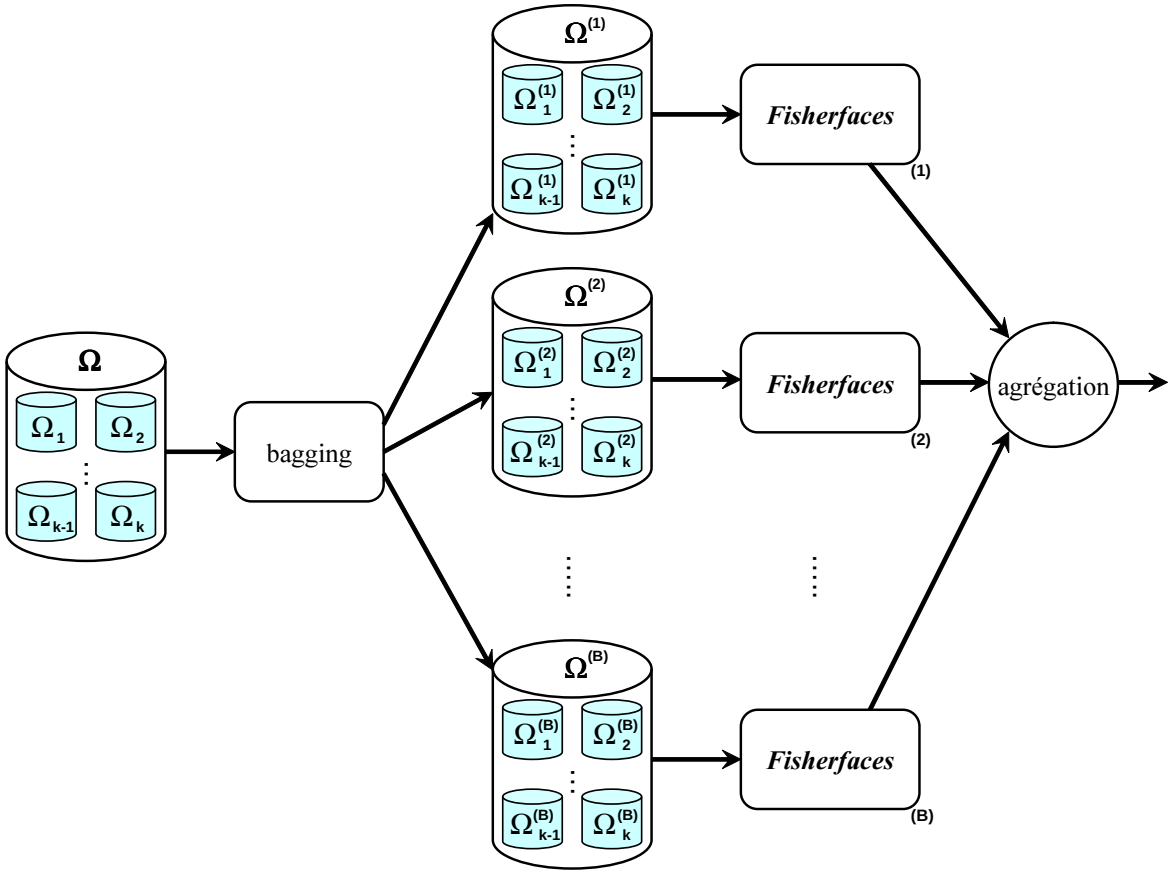


FIG. G.2 – *Technique de Lu et Jain [LJ03]. On construit B sous-échantillons à partir de la base initiale Ω, dans lesquels on n'inclut qu'un certain nombre $N'_j < N_j$ d'exemples par classe, sélectionnés aléatoirement. Puis, un classifieur fisherfaces est construit depuis chacun de ces sous-échantillons. Les résultats de classification de chacun de ces classifieurs sont agrégés selon une règle de vote à la majorité.*

3. on construit B_1 classifieurs *fisherfaces* depuis la projection de la base d'apprentissage entière sur chacun des sous-espaces $P^{(b)}$.

Lorsqu'un visage-requête doit être reconnu, il est projeté simultanément dans chacun des sous-espaces d'eigenfaces $P^{(b)}$, puis chaque classifieur ADL lui assigne une identité. Wang et Tang préconisent l'utilisation d'un simple vote à la majorité pour fusionner les résultats de classifications des B_1 classifieurs. Le choix de garder les N_0 premières *eigenfaces* dans tous les sous-espaces vise à pourvoir tous les classifieurs ADL d'une somme d'information structurelle minimale ; sans cela (c'est-à-dire si $N_0 = 0$), Wang et Tang ont montré par le biais d'expérimentations (menées sur la base XM2VTS [MMK+99]) que les résultats de classification de la plupart des B_1 classifieurs seraient très faibles, diminuant ainsi les performances du système. Le fait que les N_1 dernières *eigenfaces* soient choisies aléatoirement assure une certaine diversité dans les classifieurs. La technique de rééchantillonnage utilisée étant proche de la technique des sous-espaces aléatoires, le classifieur global ainsi construit a un comportement proche de celui d'un classifieur *fisherfaces* issu de plus d'exemples. La phase de rééchantillonnage permet donc de sortir de la zone où la taille de la base est critique, et d'améliorer les capacités de généralisation des

fisherfaces. Il faut noter également que cette technique ne rejette *a priori* aucune information hors du noyau de S_T, qui ne contient pas d'information discriminante [HLLM02]. Il semblerait que l'utilisation de $B_1 = 20$ sous-espaces donne de bons résultats. Néanmoins, aucune stratégie de choix des paramètres N_0^B et N_1^B, ni même de la taille des sous-espaces $N_0^B + N_1^B$, n'est fournie.

La deuxième technique proposée par Wang et Tang allie l'ADL_0 (*cf.* section F.1.1 de l'annexe F), et une technique proche du *bagging* (voir figure G.3). Rappelons que la technique d'ADL_0 ne retient aucune information hors du noyau de S_w. Selon Cevikalp *et al.* [CNWB05], plus la taille du noyau est importante, plus l'ADL_0 est performante. Dans le but d'encourager ce phénomène, Wang et Tang choisissent de la coupler avec une technique proche du *bagging*. Chaque classifieur est construit à partir de moins de données indépendantes et la taille du noyau de sa matrice de covariance intra-classe est augmentée. Afin d'éviter de manipuler directement la matrice de covariance intra-classe des vecteurs-images, de très grandes dimensions, Wang et Tang appliquent une ACP en amont de la phase d'ADL, à la manière de Huang *et al.* [HLLM02], de manière à neutraliser uniquement le noyau de S_T. Pour ne pas avoir à construire une ACP par sous-échantillon, l'ACP est effectuée en amont du *bagging*. La base d'apprentissage est projetée sur son sous-espace principal, pour donner naissance à une nouvelle base Ω'. Puis, cette base projetée est rééchantillonnée pour obtenir B_2 sous-échantillons. Wang et Tang considèrent que, trop souvent, on ne dispose pas de suffisamment d'images par classe pour retirer des exemples des classes, comme le font Lu et Jain [LJ03]. Dans leur méthode, chaque échantillon contient un nombre $k' < k$ de classes de la base d'apprentissage, qui lui sont aléatoirement attribuées. Ainsi, le rééchantillonnage s'effectue sur les classes (chaque classe étant considérée comme une observation), et non à l'intérieur de celles-ci. Chaque classe doit être incluse dans au moins un échantillon. Un classifieur ADL_0 est construit pour chacun de ces échantillons. Les résultats de classification sont fusionnés à l'aide d'une règle de vote à la majorité. Le nombre de classes dans chacun des sous-échantillons est inférieur au nombre de classes initialement enregistrées dans la base d'apprentissage. Chaque classifieur ne pourra donc assigner une identité qu'à un nombre limité de personnes. Il faudra donc que chaque classifieur ADL_0 soit capable de rejeter un visage-requête comme non enregistré dans sa base, et cela automatiquement, afin d'éviter de bruiter le vote à la majorité. Or, Wang et Tang ne fournissent aucune méthodologie bien définie pour procéder à ce rejet.

Wang et Tang considèrent que les deux techniques présentées ci-dessus sont complémentaires (au sens où les modèles de *fisherfaces* sont construits dans l'espace-image de S_w, tandis que les classifieurs ADL_0 sont construits depuis le noyau de S_w) et qu'elles peuvent donc être fusionnées. Le schéma de fusion est présenté en figure G.3. Les $B_1 + B_2$ classifieurs issus de ces deux techniques sont agrégés par l'utilisation d'une règle de vote à la majorité ou d'une règle de la somme. Les résultats des expériences menées par Wang et Tang sur la base XM2VTS [MMK+99] montrent que la première méthode semble être légèrement plus performante que la seconde, et que la fusion des deux techniques améliore les résultats de classification par rapport à chacune d'elles, prise séparément.

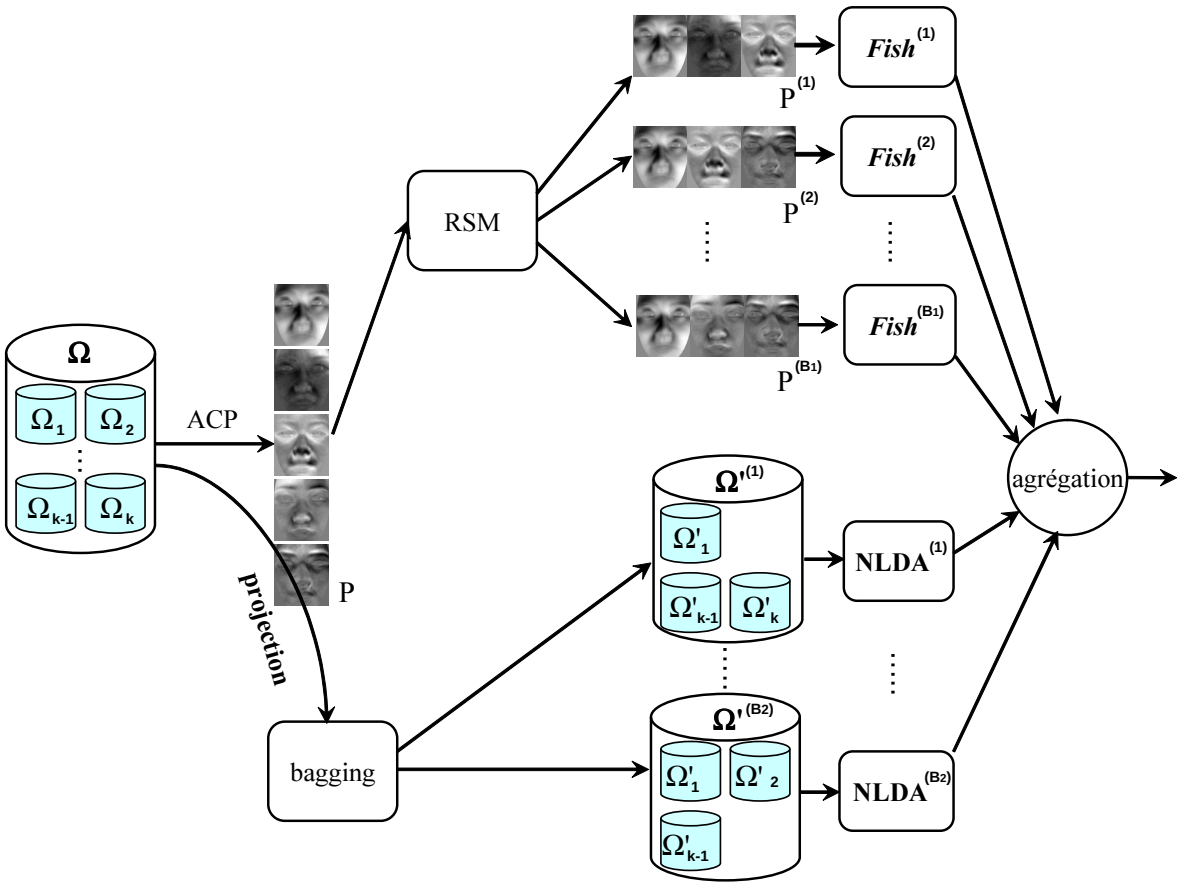


FIG. G.3 – *Agrégation des techniques de ACP+SEA+fisherfaces et des techniques d'ACP+bagging+ADL*$_0$.

Bibliographie

[ABR64] M.A. Aizerman, E.M. Braverman et L.I. Rozonoér. « Theoretical Fundations of the Potential Function Method in Pattern Recognition Learning ». *Automatic and Remote Control*, 25:821–837, 1964.

[AHK01] C.C. Aggarwal, A. Hinneburg et D.A. Keim. « On the Surprising Behavior of Distance Metrics in High Dimensional Spaces ». Dans *Proceedings of the 8th International Conference on Database Theory*, pages 420–434, 2001.

[Ahl01] J. Ahlberg. « Using the Active Appearance Algorithm for Face and Facial Feature Tracking ». Dans *Proceedings of the IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS'01)*, pages 68–71, 2001.

[AL77] S.W. Ahmed et P.A. Lachenbruch. « Discriminant Analysis when Scale Contamination is Present in the Initial Sample ». Dans J. Van Rysin, éditeur, *Classification and Clustering*, pages 331–353. Academic Press, New York, N.Y., 1977.

[Bar81] R. Baron. « Mechanisms of Human Facial Recognition ». *International Journal of Man Machine Studies*, 15:136–178, 1981.

[Bau72] L.E. Baum. « An Inequality and Associated Maximisation Technique in Statistical Estimation for Probabilistic Functions of Markov Processes ». *Inequalities*, 3:1–8, 1972.

[BBP01] D.M. Blackburn, M. Bone et P.J. Philips. « Facial Recognition Vendor Test 2000: Evaluation Report ». Technical Report A269514, National Institute of Standards and Technology, 2001. 70 pages.

[BBTD03] R. Beveridge, D. Bolme, M. Teixeira et Bruce Draper. « The CSU Face Identification Evaluation System Users Guide: Version 5.0 ». Technical report, Computer Science Department, Colorado State University, 2003. 29 pages.

[BDBS02] K. Baek, B. Draper, J.R. Beveridge et K. She. « PCA vs. ICA: a Comparison on the FERET Data Set ». Dans *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 824–827, 2002.

[Bey94] D. Beymer. « Face Recognition Under Varying Pose ». Dans *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 756–761, 1994.

[BG94] G. Bors et M. Gabbouj. « Minimal Topology for a Radial Basis Functions Neural Networks for Pattern Classification ». *Digital Processing*, 4:173–188, 1994.

[BGS05] C. Bouveyron, S. Girard et C. Schmid. « Analyse Discriminante de Haute Dimension ». Rapport de recherche n°5470, INRIA, Janvier 2005. 43 pages.

[BHB98] V. Bruce, P.J.B. Hancock et A.M. Burton. « Human Face Perception and Identification ». Dans H. Wechsler, P.J. Phillips, V. Bruce, F.F. Soulie

et T.S. Huang, éditeurs, *Face Recognition: from Theory to Applications*, pages 51–72. Springer-Verlag, Berlin, Allemagne, 1998.

[BHK97] P.N. Belhumeur, J.P. Hespanha et D.J. Kriegman. « Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[BJ02] F.R. Bach et M.I. Jordan. « Kernel Independent Component Analysis ». *Journal of Machine Learning Research*, 3:1–48, 2002.

[BJL03] S. Bres, J.M. Jolion et F. Lebourgeois. *Traitement et analyse des images numériques*. Hermes, 2003. 412 pages.

[BJS86] I.O. Bohachevsky, M.E. Johnson et M.L. Stein. « Generalized Simulated Annealing for Function Optimization ». *Technometrics*, 28(3):209–217, 1986.

[BK98] I. Biederman et P. Kalocsai. « Neural and Psychophysical Analysis of Object and Face Recognition ». Dans H. Wechsler, P.J. Phillips, V. Bruce, F.F. Soulie et T.S. Huang, éditeurs, *Face Recognition: from Theory to Applications*, pages 3–25. Springer-Verlag, Berlin, Allemagne, 1998.

[BL88] D.S. Broomhead et D. Lowe. « Multivariable Functional Interpolation and Adaptive Networks ». *Complex Systems*, 2:321–355, 1988.

[BL94] V. Barnett et T. Lewis. *Outliers in Statistical Data*. John Wiley, New York, N.Y., 1994. 604 pages.

[Ble66] W.W. Bledsoe. « The Model Method in Facial Recognition ». Technical Report PRI:15, Panoramic Research Inc., Palo Alto, CA, 1966.

[BLFM03] M.S. Bartlett, G. Littlewort, I. Fasel et J.R. Movellan. « Real Time Recognition of Facial Expressions: Development and Applications to Human Computer Interaction ». Dans *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 5, pages 53–58, 2003.

[BLP02] Y. Bing, J. Lianfu et C. Ping. « A New LDA-based Method for Face Recognition ». Dans *Proceedings of the 16th International Conference on Pattern Recognition (ICPR02)*, volume 1, pages 168–171, 2002.

[BM95] R. Brunelli et S. Messelodi. « Robust Estimation of Correlation with Applications to Computer Vision ». *Pattern Recognition*, 28:833–841, 1995.

[BMS02] M.S. Bartlett, J.R. Movellan et T.J. Sejnowski. « Face Recognition by Independent Component Analysis ». *IEEE Transactions on Neural Networks*, 13(6):1450–1464, 2002.

[BP93] R. Brunelli et T. Poggio. « Face Recognition: Features versus Templates ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, 1993.

[BP94] M. Bichsel et A. Pentland. « Human Face Recognition and Face Image Set's Topology ». *CVGIP: Image Understanding*, 59:254–261, 1994.

[Bre96] L. Breiman. « Bagging Predictors ». *Machine Learning*, 24(2):123–140, 1996.

[BS95] A. Bell et T. Sejnowski. « An Information Maximization Approach to Blind Separation and Blind Deconvolution ». *Neural Computation*, 7:1129–1159, 1995.

[BSDG01] J.R. Beveridge, K. She, B.A. Draper et G.H. Givens. « A Nonparametric Statistical Comparison of Principal Component and Linear Discriminant Subspaces for Face Recognition ». Dans *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 535–542, 2001.

[BT87] N. Balakrishnan et M.L. Tiku. « Robust Classification Procedures ». Dans *Proceedings of the International Conference of the International Federation of Classification Societies*, pages 269–276, 1987.

[Bug98] G. Bugmann. « Classification using Networks of Normalized Radial Basis Functions ». Dans *Proceedings of the International Conference on Advances in Pattern Recognition ICAPR'98*, pages 435–444, 1998.

[Bur88] P. Burt. « Smart Sensing Within a Pyramid Vision Machine ». *Proceedings of the IEEE*, 76:1006–1015, 1988.

[But94] L.J. Buturovic. « Toward Bayes-Optimal Linear Dimension Reduction ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(4):420–424, 1994.

[BVB87] V. Bruce, T. Valentine et A. Baddeley. « The Basis of the 3/4 Face Advantage in Face Recognition ». *Applied Cognitive Psychology*, pages 109–120, 1987.

[BY98] V. Bruce et A. Young. *In the Eye of the Beholder: The science of Face Perception*. Oxford University Press, 1998. 280 pages.

[Cam82] N.A. Campbell. « Robust Procedures in Multivariate Analysis II: Robust Canonical Variate Analysis ». *Applied Statistics*, 31:1–8, 1982.

[Car99] J.F. Cardoso. « Higher-order Contrasts for Independent Component Analysis ». *Neural Computation*, 11(1):157–192, 1999.

[CBF05] K.I. Chang, K.W. Bowyer et P.J. Flynn. « An Evaluation of Multi-Modal 2D+3D Face Biometrics ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):619–624, Avril 2005.

[CC04] D. Cristinacce et T. Cootes. « A Comparison of Shape Constrained Facial Feature Detectors ». Dans *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 375–380, 2004.

[CD01] C. Croux et C. Dehon. « Robust Linear Discriminant Analysis using S-estimators ». *The Canadian Journal of Statistics*, 29:473–492, 2001.

[CET01] T.F. Cootes, G.J. Edwards et C.J. Taylor. « Active Appearance Models ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[CF90] G. Cotrell et M. Fleming. « Face Recognition using Unsupervised Feature Extraction ». Dans *Proceedings of the International Conference on Neural Network*, pages 322–325, 1990.

[CH97] A. Colmenarez et T. Huang. « Face Detection with Information-Based Maximum Discrimination ». Dans *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 782–787, 1997.

[CLK+00] L. Chen, H. Liao, M. Ko, J. Lin et G. Yu. « A New LDA-based Face Recognition System Which can Solve the Small Sample Size Problem ». *Pattern Recognition*, 33(10):1713–1726, 2000.

[CLLC04] J. Cheng, Q. Liu, H. Lu et Y.W. Chen. « Random Independent Subspace for Face Recognition ». Dans *Proceedings of the International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES 2004)*, pages 352–358, 2004.

[CLLH01] L.F. Chen, H.Y.M. Liao, J.C. Lin et C.C. Han. « Why Recognition in a Statistics-based Face Recognition System Should be Based on the Pure Face Portion: a Probabilistic Decision-Based Proof ». *Pattern Recognition*, 34(7):1393–1403, 2001.

[CNWB05] H. Cevikalp, M. Neamtu, M. Wilkes et A. Barkana. « Discriminative Common Vectors for Face Recognition ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):4–13, Janvier 2005.

[Com94] P. Comon. « Independent Component Analysis, a new concept ? ». *Signal Processing, Special issue on Higher-Order Statistics.*, 36(3):287–314, 1994.

[Cov65] T.M. Cover. « Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition ». *IEEE Transactions on Electronic Computers*, 14:326–334, 1965.

[CR92] C.V. Chork et P.J. Rousseeuw. « Integrating a High-Breakdown Option into Discriminant Analysis in Exploration Geochemistry ». *Journal of Geochemical Exploration*, 43:191–203, 1992.

[CYHD05] W.S. Chen, P.C. Yuen, J. Huang et D.Q. Dai. « Kernel Machine-Based One-Parameter Regularized Fisher Discriminant Method for Face Recognition ». *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 35(4):659–669, Août 2005.

[DAD03] F. Davoine, B. Abboud et M. Dang. « Analyse de Visages et d'Expressions Faciales par Modèle Actif d'Apparence ». *Traitement du Signal*, 21(3):179–193, 2003.

[DBBB03] B.A. Draper, K. Baek, M.S. Bartlett et J.R. Beveridge. « Recognizing Faces with PCA and ICA ». *Computer Vision and Image Understanding*, 91(1-2):115–137, 2003.

[DC86] R. Diamon et S. Carey. « Why Faces are and are not Special: An effect of Expertise ». *Journal of Experimental Psychology*, 115(2):107–117, 1986.

[DG05] S. Duffner et C. Garcia. « A Connexionist Approach for Robust and Precise Facial Feature Detection in Complex Scenes ». Dans *Proceedings of the 4th IEEE Symposium on Image and Signal Processing and Analysis (ISPA 2005)*, Septembre 2005. À paraître.

[DGG05] K. Delac, M. Grgic et S. Grgic. « A Comparative Study of PCA, ICA and LDA ». Dans *Proceedings of the 5th EURASIP Conference on Speech and Image Processing*, pages 99–106, Juillet 2005.

[DHS01] R.O. Duda, P.E. Hart et D.G. Stork. *Pattern Classification.* Wiley-Interscience, 2001. 680 pages.

[DJK+02] J.L. Dugelay, J.C. Junqua, C. Kotropoulos, R. Kuhn, F. Perronnin et I. Pitas. « Recent Advances in Biometric Person Authentication ». Dans *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP 2002)*, volume 4, pages 4060–4063, 2002.

[DK82] P.A. Devijver et J. Kittler. *Pattern Recognition: A Statistical Approach.* Prentice-Hall, London, 1982.

[DM77] H.P. Decell et S.M. Mayekar. « Feature Combinations and the Divergence Criterion ». *Computers and Mathematics with Applications*, 3:71–76, 1977.

[DQJ04] G. Dai, Y. Qian et S. Jia. « A Kernel Fractional-Step Nonlinear Discriminant Analysis for Pattern Recognition ». Dans *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*, pages 431–434, 2004.

[Dui95] R.P.W. Duin. « Small Sample Size Generalization ». Dans *Proceedings of the 9th Scandinavian Conference on Image Analysis*, volume 2, pages 957–964, 1995.

[Dui00] R.P.W. Duin. « Classifiers in Almost Empty Spaces ». Dans *Proceedings of the 15th International Conference on Pattern Recognition (ICPR'00)*, pages 1–7, 2000.

[DY03] D.Q. Dai et P.C. Yuen. « Regularized Discriminant Analysis and its Application to Face Recognition ». *Pattern Recognition*, 36(3):845–847, 2003.

[ECT98] G.J. Edwards, T.F. Cootes et C.J. Taylor. « Face Recognition Using Active Appearance Models ». Dans *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 2, pages 581–595, 1998.

[ET93] B. Efron et R.J. Tibshirani. *An Introduction To The Bootstrap*. Monographs on Statistics and Applied Probability 57, Chapman & Hall, New York, 1993. 440 pages.

[ETC98] G.J. Edwards, C.J. Taylor et T.F. Cootes. « Interpreting Face Images Using Active Appearance Models ». Dans *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 300–305, 1998.

[EWLT02] M.J. Er, S. Wu, J. Lu et H.L. Toh. « Face Recognition With Radial Basis Function (RBF) Neural Networks ». *IEEE Transactions on Neural Networks*, 13:697–709, 2002.

[FBVC01] R. Féraud, O.J. Bernier, J.-E. Viallet et M. Collobert. « A Fast and Accurate Face Detection Based on Neural Network ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(1):42–53, 2001.

[Fis36] R.A. Fisher. « The Use of Multiple Measures in Taxonomic Problems ». *Annals of Eugenics*, 17:179–188, 1936.

[FL03a] B. Fasel et J. Luettin. « Automatic Facial Expression Analysis: A Survey ». *Pattern Recognition*, 36:259–275, 2003.

[FL03b] S. Fidler et A. Leonardis. « Robust LDA Classification by Subsampling ». Dans *IEEE Workshop on Statistical Analysis in Computer Vision*, volume 8, pages 97–104, 2003.

[For73] G.D. Forney. « The Viterbi Algorithm ». *Proceedings of the IEEE*, 61(3):268–278, 1973.

[Fri89] J.H. Friedman. « Regularized Discriminant Analysis ». *Journal of the American Statistical Association*, 84:165–175, 1989.

[FS96] Y. Freund et R.E. Schapire. « Experiments with a New Boosting Algorithm ». Dans *Proceedings of the 13th International Conference on Machine Learning (ICML'96)*, pages 148–156, 1996.

[FS97] Y. Freund et R.E. Schapire. « A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting ». *Journal of Computer and System Sciences*, 55:119–139, 1997.

[FTV99] R.Q. Feitosa, C.E. Thomaz et A. Veiga. « Comparing the Performance of the Discriminant Analysis and RBF Neural Network for Face Recognition ». Dans *Proceedings of International Conference on Information Systems Analysis and Synthesis (ISAS'99)*, volume 6, 1999. 8 pages.

[Fuk90] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, Inc, seconde édition, 1990. 369 pages.

[GA98] D. Graham et N. Allison. « Characterizing Virtual Eigensignatures for General Purpose Face Recognition ». Dans H. Wechsler, P.J. Phillips, V. Bruce, F.F.

Soulie et T.S. Huang, éditeurs, *Face Recognition: from Theory to Applications*, pages 446–456. Springer-Verlag, Berlin, Allemagne, 1998.

[GAP+02] H. Gupta, A.K. Agrawal, T. Pruthi, C. Shekhar et R. Chellappa. « An Experimental Evaluation of Linear and Kernel-Based Methods for Face Recognition ». Dans *Proceedings of the IEEE Workshop on Application of Computer Vision (WACV) 2002*, pages 13–19, 2002.

[GBM03] « The Global Biometrics Market ». Report G-276, Business Communications Company, Inc, Norwalk, CT, 2003.

[GD04] C. Garcia et M. Delakis. « Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11):1408–1423, 2004.

[GL02] Y. Gao et K.H. Leung. « Face Recognition using Line Edge Map ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):764–779, 2002.

[GMP00] S. Gong, S.J. McKenna et A. Psarrou. *Dynamic Vision: From Images to Face Recognition*. Imperial College Press, Londres, Angleterre, 2000. 364 pages.

[GN00] J. Ghosh et A. Nag. « An Overview of Radial Basis Function Networks ». Dans R.J. Howlerr et L.C. Jain, éditeurs, *Radial Basis Function Neural Network Theory and Applications*. Physica-Verlag, 2000. 28 pages.

[Gow66] J.C. Gower. « Some Distance Properties of Latent Root and Vector Methods used in Multivariate Analysis ». *Biometrika*, 53:325–338, 1966.

[Gro04] R. Gross. « Face Databases » . Dans S.Z. Li et A.K. Jain, éditeurs, *Handbook of Face Recognition*, Chapitre 13. Springer-Verlag, Reidel, Dordrecht, 2004. 22 pages.

[GSC01] R. Gross, J. Shi et J.F. Cohn. « Quo Vadis Face Recognition? The Current State of the Art in Face Recognition ». Technical Report CMU-RI-TR-01-17, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 2001. 25 pages.

[GT97] I. Gauthier et M.J. Tarr. « Becoming a "Greeble" expert: Exploring the Face Recognition Mechanism ». *Vision Research*, 37(12):1673–1682, 1997.

[GT00] C. Garcia et G. Tziritas. « Wavelet Packet Analysis for Face Recognition ». *Image and Vision Computing*, 18(4):289–297, 2000.

[Hay94] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Upper Saddle River, NJ, USA, 1994. 768 pages.

[HB96] H. Hill et V. Bruce. « Effects of Lighting on Matching Facial Surfaces ». *Journal of Experimental Psychology: Human Perception and Performance*, 22:986–1004, 1996.

[HF00] X. He et W. K. Fung. « High Breakdown Estimation for Multiple Populations with Applications to Discriminant Analysis ». *Journal of Multivariate Analysis*, 72:151–162, 2000.

[HJP03] P. Howland, M. Jeon et H. Park. « Structure Preserving Dimension Reduction for Clustered Text Data Based on the Generalized Singular Value Decomposition ». *SIAM Journal on Matrix Analysis and Applications*, 25:165–179, 2003.

[HKR93] D.P. Huttenlocher, G.A. Klanderman et W.A. Rucklidge. « Comparing Images Using the Hausdorff Distance ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993.

[HL01] E. Hjelmas et B.K. Low. « Face Detection: A Survey ». *Computer Vision and Image Understanding*, 83:236–274, 2001.

[HLLM02] R. Huang, Q. Liu, H. Lu et S. Ma. « Solving the Small Sample Size Problem of LDA ». Dans *Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02)*, volume 3, pages 29–32, 2002.

[HM97] D. M. Hawkins et G. J. McLachlan. « High-breakdown Linear Discriminant Analysis ». *Journal of the American Statistical Association*, 92(437):136–143, 1997.

[HO97] A. Hyvärinen et E. Oja. « A Fast fixed-point algorithm for Independent Component Analysis ». *Neural Computation*, 9(7):1483–1492, 1997.

[Ho98] T.K. Ho. « The Random Subspace Method for Constructing Decision Forests ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:832–844, 1998.

[Hot33] H.J. Hotelling. « Analysis of Statistical Variables into Principal Components ». *Journal of Educational Psychology*, 24:417–441, 1933.

[HP04] P. Howland et H. Park. « Generalizing Discriminant Analysis Using the Generalized Singular Value Decomposition ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8):995–1006, 2004.

[HRL04] B.W. Hwang, M.C. Roh et S.W. Lee. « Performance Evaluation of Face Recognition Algorithms on Asian Face Database ». Dans *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 278–283, 2004.

[HS89] T. Hastie et W. Stuetzle. « Principal Curves ». *Journal of the American Statistical Association*, 84:502–516, 1989.

[HSA97] H. Hill, P.G. Schyns et S. Akamatsu. « Information and Viewpoint Dependence in Face Recognition ». *Cognition*, 62:201–222, 1997.

[HT96] T. Hastie et R. Tibshirani. « Discriminant Analysis by Gaussian Mixtures ». *Journal of the Royal Statistical Society series B*, 58:158–176, 1996.

[Hub85] P.J. Huber. « Projection Pursuit ». *The Annals of Statistics*, 13:435–525, 1985.

[HVD00] M. Hubert et K. Van Driessen. « Fast and Robust Discriminant Analysis ». Technical Report 2002-09 , Katholieke Universiteit Leuven, Department of Mathematics, University Centre for Statistics, 2000. 22 pages.

[HYCL03] J. Huang, P.C. Yuen, W.S. Chen et J.H. Lai. « Component-based LDA Method For Face Recognition with One Training Sample ». Dans *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pages 120–126, 2003.

[Hyv99] A. Hyvärinen. « Survey on Independant Component Analysis ». *Neural Computing Surveys*, 2:94–128, 1999.

[JC82] A.K. Jain et B. Chandrasekaran. « Dimensionality and Sample Size Considerations in Pattern Recognition Practice » . Dans *Handbook of Statistics 2*, pages 835–855. Krishnaiah, P.R. and Kanal, L.N., Amsterdam, Pays-Bas, 1982.

[JD88] A.K. Jain et R.C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall Advanced Reference Series, Upper Saddle River, New Jersey, USA, 1988. 320 pages.

[JH91] C. Jutten et J. Herault. « Blind Separation of Sources ». *Signal Processing*, 24:1–10, 1991.

[JHC92] A. Johnston, H. Hill et N. Carman. « Recognizing Faces: Effects of Lighting Direction, Inversion and Brightness Reversal ». *Cognition*, 40:1–19, 1992.

[JKLM00] K. Jonsson, J. Kittler, Y. Li et J. Matas. « Learning Support Vectors for Face Verification and Recognition ». Dans *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 208–213, 2000.

[Jol86] I.T. Joliffe. *Principal Component Analysis*. Springer-Verlag, New York, 1986.

[Jol01] J.M. Jolion. « Graph Matching: What are we Talking About? ». Dans *Proceedings of the IAPR Workshop on Graph-based Representations in Pattern Recognition*, pages 170–175, 2001.

[JS93] J.S.R. Jang et C.T. Sun. « Functional Equivalence Between Radial Basis Function Networks and Fuzzy Inference Systems ». *IEEE Transactions on Neural Networks*, 4:156–159, 1993.

[KA98] N. Kumar et A.G. Andreou. « Heteroscedastic Discriminant Analysis and Reduced Rank HMMs for Improved Speech Recognition ». *Speech Communication*, 26:283–297, 1998.

[KAG03] N. Kumar, V. Abhishek et G. Gautam. « A Novel Approach for Person Authentication and Content-Based Tracking in Videos using Kernel Methods and Active Appearance Models ». Dans *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, volume 2, pages 1384–1389, 2003.

[Kan77] T. Kanade. *Computer Recognition of Human Faces*. Interdisciplinary Systems Research, 1977. 106 pages.

[KCT00] T. Kanade, J.F. Cohn et Y.L. Tian. « Comprehensive Database for Facial Expression Analysis ». Dans *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 46–53, 2000.

[Kel70] M.D. Kelly. « Visual Identification of People by Computer ». Technical Report AI-130, Stanford AI Project, Stanford, CA, 1970. 247 pages.

[KHDM98] J. Kittler, M. Hatef, R.P.W. Duin et J. Matas. « On Combining Classifiers ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.

[Kir00] M. Kirby. *Dimensionality Reduction and Pattern Analysis: An Empirical Approach*. Wiley, New York, 2000.

[KJMT95] W.J. Krzanowski, P. Jonathan, W.V. McCarthy et M.R. Thomas. « Discriminant Analysis with Singular Covariance Matrices: Methods and Applications to Spectroscopic Data ». *Applied Statistics*, 44:101–115, 1995.

[KKB02] H.C. Kim, D. Kim et S.Y. Bang. « Face Recognition using the Mixture-of-Eigenfaces Method ». *Pattern Recognition Letters*, 23(13):1549–1558, 2002.

[Koh89] T. Kohonen. *Self-Organizing and Associative Memory*. Springer-Verlag, Berlin, 1989.

[Krü97] N. Krüger. « An Algorithm for the Learning of Weights in Discrimination Functions Using a Priori Constraints ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):764–768, 1997.

[Kra91] M.A. Kramer. « Nonlinear Principal Component Analysis using Autoassociative Neural Networks ». *AIChe Journal*, 32:233–243, 1991.

[KS90] M. Kirby et L. Sirovich. « Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, 1990.

[KTP00] C.L. Kotropoulos, A. Tefas et I. Pitas. « Frontal Face Authentication using Discriminating Grids with Morphological Feature Vectors ». *IEEE Transactions on Multimedia*, 2:14–26, 2000.

[LCBD+90] Y. Le Cun, B. Boser, J.S. Denker, D. Henderson, R. Howard, W. Hubbard et L. Jackel. « Handwritten Digit Recognition with a Backpropagation Neural Network ». Dans *Proceedings of the International Conference on Advances in Neural Information Processing Systems (NIPS)*, volume 2, pages 396–404, 1990.

[LD04] M. Loog et R.PW. Duin. « Linear Dimensionality Reduction via a Heteroscedastic Extension of LDA: the Chernoff Criterion ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):732–739, 2004.

[LDGF04] S. Lespinats, P. Deschavanne, A. Giron et B. Fertil. « Pertinence des Métriques Fractionnaires pour l'Analyse des Données de Grande Dimension ». Dans *Actes des Journées d'Extraction et de Gestion des Connaissances*, pages 135–142, 2004.

[LDHU01] M. Loog, R.P.W. Duin et R. Haeb-Umbach. « Multiclass Linear Dimension Reduction by Weighted Pairwise Fisher Criteria ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(7):762–766, 2001.

[LGTB97] S. Lawrence, C. Giles, A. Tsoi et A. Black. « Face Recognition: A Convolutional Neural-Network Approach ». *IEEE Transactions on Neural Networks*, 8:98–112, 1997.

[LHLM02] Q. Liu, R. Huang, H. Lu et S. Ma. « Face Recognition Using Kernel Based Fisher Discriminant Analysis ». Dans *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 197–207, 2002.

[Liu03] C. Liu. « A Bayesian Discriminating Features Method for Face Detection ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):725–740, 2003.

[LJ03] X. Lu et A.K. Jain. « Resampling for Face Recognition ». Dans *Proceedings of the 4th International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA 2003)*, pages 869–877, 2003.

[LK00] R. Lotlikar et R. Kothari. « Fractional-Step Dimensionality Reduction ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(6):623–627, 2000.

[LKL97] S.H. Lin, S.Y. Kung et L.J. Lin. « Face Recognition/Detection by Probabilistic Decision-based Neural Networks ». *IEEE Transactions on Neural Networks*, 8:114–132, 1997.

[LL99] S.Z. Li et J. Lu. « Face Recognition Using the Nearest Feature Line Method ». *IEEE Transactions on Neural Networks*, 10:439–443, 1999.

[LLM04] Q. Liu, , H. Lu et S. Ma. « Improving Kernel Fisher Discriminant Analysis for Face Recognition ». *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):42–49, 2004.

[Loè55] M.M. Loève. *Probability Theory.* Van Nostrand, Princeton, N.J., 1955. 297 pages.

[Loo00] M. Loog. *Approximate Pairwise Accuracy Criteria for Multiclass Linear Dimension Reduction - Generalisations of the Fisher Criterion.* Delft University Press, 2000. 74 pages.

[LP92] E. Levin et R. Pieraccini. « Dynamic Planar Warping for Optical Charac-

ter Recognition ». Dans *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 149–152, 1992.

[LPV03a] J. Lu, K.N. Plataniotis et A.N. Venetsanopoulos. « Face Recognition Using Kernel Direct Discriminant Analysis Algorithms ». *IEEE Transactions on Neural Networks*, 14(1):117–126, 2003.

[LPV03b] J. Lu, K.N. Plataniotis et A.N. Venetsanopoulos. « Face Recognition using LDA-based Algorithms ». *IEEE Transactions on Neural Networks*, 14:195–200, 2003.

[LPV05] J. Lu, K.N. Plataniotis et A.N. Venetsanopoulos. « Regularization Studies of Linear Discriminant Analysis in Small Sample Size Scenarios with Application to Face Recognition ». *Pattern Recognition Letters*, 26:181–191, Janvier 2005.

[LSG04] X. Liu, A. Srivastava et K. Gallivan. « Optimal Linear Representations of Images for Object Recognition ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):662–666, 2004.

[LTC95] A. Lanitis, C.J. Taylor et T.F. Cootes. « Automatic Face Identification System Using Flexible Appearance Models ». *Image and Vision Computing*, 13:393–401, 1995.

[LVB+93] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Wurtz et W. Konen. « Distortion Invariant Object Recognition in the Dynamic Link Architecture ». *IEEE Transactions on Computers*, 42(3):300–311, 1993.

[LW98] C. Liu et H. Wechsler. « Enhanced Fisher Linear Discriminant Models for Face Recognition ». Dans *Proceedings of the International Conference on Pattern Recognition (ICPR98)*, pages 1368–1372, 1998.

[LW00] C. Liu et H. Wechsler. « Robust Coding Scheme for Indexing and Retrieval from Large Face Databases ». *IEEE Transactions on Image Processing*, 9:132–137, 2000.

[LW01] C. Liu et H. Wechsler. « A Shape- and Texture-based Enhanced Fisher Classifier for Face Recognition ». *IEEE Transactions on Image Processing*, 10:598–608, 2001.

[LWLT04] W. Liu, Y. Wang, S.Z. Li et Tan T.. « Null Space-based Kernel Fisher Discriminant Analysis for Face Recognition ». Dans *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 369–374, 2004.

[Mah36] P.C. Mahalanobis. « On the Generalized Distance in Statistics ». Dans *Proceedings of the National Institute of Sciences of India*, volume 12, pages 49–55, 1936.

[MAU94] Y. Moses, Y. Adini et S. Ullman. « Face Recognition: the Problem of Compensating for Changes in Illumination Direction ». Dans *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 1, pages 286–296, 1994.

[MB98] A.R. Martinez et R. Benavente. « The AR Face Database ». Technical Report 24, Computer Vision Center (CVC), Barcelone, Espagne, 1998.

[McL04] G.J. McLachlan. *Discriminant Analysis and Statistical Pattern Recognition.* John Wiley & Sons, 2004. 526 pages.

[MD89] J. Moody et J. Darken. « Fast Learning in Networks of Locally-Tuned Processing Unit ». *Neural Computing*, 1:281–294, 1989.

[MH98] M. Meila et D. Heckerman. « An Experimental Comparison of Several Clustering and Initialization Methods ». Dans *Proceedings of the Conference on Uncertainty in Artificial Intelligence UAI'98*, pages 386–395, 1998.

[MJP00] B. Moghaddam, T. Jebara et A. Pentland. « Bayesian Face Recognition ». *Pattern Recognition*, 33(11):1771–1782, 2000.

[MK01] A.M. Martinez et A.C. Kak. « PCA versus LDA ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):228–233, 2001.

[MMK+99] K. Messer, J. Matas, J. Kittler, J. Luettin et G. Maitre. « XM2VTSDB: The Extended M2VTS Database ». Dans *Proceedings of the 2nd International Conference on Audio- and Video Based Biometric Person Authentication (AVBPA)*, pages 72–77, 1999.

[MMR+01] K.R. Müller, S. Mika, G. Raetsch, K. Tsuda et B Schölkopf. « An Introduction to Kernel-Based Learning Algorithms ». *IEEE Transactions on Neural Networks*, 12(2):181–201, 2001.

[MN95] H. Murase et S.K. Nayar. « Visual Learning and Recognition of 3D Objects from Appearance ». *International Journal of Computer Vision*, 14:5–24, 1995.

[Mog02] B. Moghaddam. « Principal Manifolds and Probabilistic Subspaces for Visual Recognition ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):780–788, 2002.

[Moo96] T.K. Moon. « The Expectation-Maximization Algorithm ». *IEEE Signal Processing Magazine*, 13:47–60, 1996.

[Mos93] Y. Moses. « *Face Recognition. Generalization to Novel Images* ». Phd thesis, Applied Mathematics and Computer Science, The Weizmann Institute of Science, Israël, 1993.

[MP97] B. Moghaddam et A. Pentland. « Probabilistic Visual Learning for Object Representation ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:775–779, 1997.

[MP98] H. Moon et J. Phillips. « Analysis of PCA-Based Face Recognition Algorithms » . Dans K. Boyer et P.J. Phillips, éditeurs, *Empirical Evaluation Techniques in Computer Vision*, pages 835–855. IEEE Computer Society Press, 1998.

[MRW+99] S. Mika, G. Ratsch, J. Weston, B. Schölkopf et K. Müller. « Fisher Discriminant Analysis with Kernels ». Dans *Proceedings of IEEE Workshop on Neural Networks for Signal Processing*, pages 41–48, 1999.

[MSGA03] J. Ma, J.L. Sancho-Gomez et S.C. Ahalt. « Nonlinear Multiclass Discriminant Analysis ». *IEEE Signal Processing Letters*, 10(7):196–199, 2003.

[MW02] A.J. Mansfield et J.L. Wayman. « Best Practices in Testing and Reporting Performance of Biometric Devices ». NPL Report CMSC 14/02, San Jose State University, 2002. 14 pages.

[Nef96] A.V. Nefian. « *Statistical Approaches To Face Recognition* ». Qualifying examination report, School of Electrical Engineering, Georgia Institute of Technology, 1996. 32 pages.

[Nef99] A.V. Nefian. « *A Hidden Markov Model-Based Approach for Face Detection and Recognition* ». Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA., 1999.

[Nef02] A.V. Nefian. « Embedded Bayesian Networks for Face Recognition ». Dans *Pro-*

ceedings of the IEEE International Conference on Multimedia and Expo, volume 2, pages 133–136, 2002.

[Nel01] O. Nelles. *Nonlinear System Identification: from Classical Approaches to Neural Networks and Fuzzy Models*. Springer-Verlag, Berlin, 2001. 802 pages.

[OFG97] E. Osuna, R. Freund et F. Girosi. « Training Support Vector Machines: An Application to Face Detection ». Dans *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 130–136, 1997.

[PA96] P.S. Penev et J.J. Atick. « Local Feature Analysis: a General Statistical Theory for Object Representation ». *Network: Computation in Neural Systems*, 7(3):477–500, 1996.

[PCST00] J. Platt, N. Cristianini et J. Shawe-Taylor. « Large Margin DAGs for Multiclass Classification ». Dans *Proceedings of the International Conference on Advances in Neural Information Processing Systems (NIPS)*, volume 12, pages 547–553, 2000.

[Pen00] P.S. Penev. « Redundancy and Dimensionality Reduction in Sparse-Distributed Representations of Natural Objects in Terms of Their Local Features ». Dans *Proceedings of the International Conference on Advances in Neural Information Processing Systems (NIPS)*, pages 901–907, 2000.

[Per75] D. Perkins. « A Definition of Caricature and Recognition ». *Study in the Anthropology of Visual Communication*, 2:1–24, 1975.

[Per04] F. Perronnin. « *A Probabilistic Model of Face Mapping Applied to Person Recognition* ». Thèse de doctorat, École Polytechnique Fédérale de Lausanne, Faculté d'Informatique et Communications, 2004.

[PG90a] T. Poggio et F. Girosi. « Networks for Approximation and Learning ». *Proceedings of the IEEE*, 78:1481–1497, 1990.

[PG90b] T. Poggio et F. Girosi. « Regularization Algorithms for Learning that are Equivalent to Multilayer Networks ». *Science*, 247:978–982, 1990.

[PG05] J.R. Price et T.F. Gee. « Face Recognition Using Direct, Weighted Linear Discriminant Analysis and Modular Subspaces ». *Pattern Recognition*, 38(2):209–219, Janvier 2005.

[PGM+03] P.J. Phillips, P. Grother, R.J Micheals, D.M. Blackburn, E. Tabassi et J.M. Bone. « Face Recognition Vendor Test 2002. Evaluation Report. ». Technical Report 6965, National Institute of Standards and Technology, 2003. 56 pages.

[Phi99] P.J. Phillips. « Support Vector Machines applied to Face Recognition ». Dans *Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems*, pages 803–809, 1999.

[Pir01] A. M. Pires. « Robust Discriminant Analysis and the Projection Pursuit Approach ». Dans *Proceedings of the International Conference on Robust Statistics*, pages 317–329, 2001.

[Pla05] V. Planchon. « Traitement des valeurs aberrantes : concepts actuels et tendances générales ». *Biotechnologie, Agronomie, Société et Environnement*, 9(1):19–34, Mars 2005.

[PMRR00] P.J. Phillips, H. Moon, S.A. Rizvi et P.J. Rauss. « The FERET Evaluation Methodology for Face-Recognition Algorithms ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1090–1104, 2000.

[PMS94] A.D. Pentland, B. Moghaddam et T. Starner. « View-Based and Modular Eigenspaces for Face Recognition ». Dans *Proceedings of the IEEE Computer Society Conference on Pattern Recognition*, pages 84–91, 1994.

[PPP04] C.H. Park, H. Park et P. Pardalos. « A Comparative Study of Linear and Nonlinear Feature Extraction Methods ». Dans *Proceedings of the Fourth IEEE International Conference on Data Mining (ICDM'04)*, pages 495–498, 2004.

[PR00] M. Pantic et J.M. Rothcrantz. « Automatic Analysis of Facial Expressions: State of the Art ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, 2000.

[PS91] J. Park et I.W. Sandberg. « Universal Approximation Using Radial Basis Functions Network ». *Neural Computation*, 3(2):246–257, 1991.

[PS04] D. Potts et C. Sammut. « Online Nonlinear System Identification in High Dimensional Environments ». Dans *Proceedings of the Australasian Conf. on Robotics and Automation*, 2004. 9 pages.

[PWHR98] P.J. Phillips, H. Wechsler, J. Huang et P.J. Rauss. « The FERET Database and Evaluation Procedure for Face Recognition Algorithms ». *Image and Vision Computing*, 16:295–306, 1998.

[Rao48] C.R. Rao. « The Utilization of Multiple Measurements in Problems of Biological Classification ». *Journal of the Royal Statistical Society, Series B.*, 10:159–203, 1948.

[RBBV04] S. Romdhani, V. Blanz, C. Basso et T. Vetter. « Morphable Models of Faces » . Dans S.Z. Li et A.K. Jain, éditeurs, *Handbook of Face Recognition*, Chapitre 10. Springer-Verlag, Reidel, Dordrecht, 2004. 33 pages.

[RBK98] H. Rowley, S. Baluja et T. Kanade. « Neural Network-Based Face Detection ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.

[RBRH78] R.H. Randles, J.D. Broffit, J.S. Ramberg et R.V. Hogg. « Generalized Linear and Quadratic Discriminant Functions Using Robust Estimates ». *Journal of the American Statistical Association*, 73:564–568, 1978.

[RD98] S. Raudys et R.P.W. Duin. « On Expected Classification Error of the Fisher Linear Classifier with Pseudo-Inverse Covariance Matrix ». *Pattern Recognition Letters*, 19:385–392, 1998.

[RJ91] S.J. Raudys et A.K. Jain. « Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners ». *PAMI*, 13:252–264, 1991.

[RML90] G. Rhodes et I.G Mc Lean. « Distinctiveness and Expertise Effects with Homogeneous Stimuli: Towards a model of Configurational Coding ». *Perception*, 19:773–794, 1990.

[RMRG97] A.V. Rao, D. Miller, K. Rose et A. Gersho. « Mixture of experts regression modeling by deterministic annealing ». *IEEE Transactions on Signal Processing*, 45:2811–2820, 1997.

[Row97] S. Roweis. « EM Algorithms for PCA and SPCA ». Dans *Proceedings of the International Conference on Advances in Neural Information Processing Systems (NIPS)*, volume 10, pages 626–632, 1997.

[RPG99] S. Romdhani, A. Psarrou et S. Gong. « Multi-View Nonlinear Active Shape Model using Kernel PCA ». Dans *Proceedings of the 10th British Machine Vision Conference*, pages 483–492, 1999.

[RPM98] S.A. Rizvi, P.J. Phillips et H. Moon. « A Verification Protocol and Statistical Performance Analysis for Face Recognition Algorithms ». Dans *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 833–838, 1998.

[RW99] M.C. Röhl et C. Weihs. « Optimal vs. Classical Linear Dimension Reduction ». Dans *Proceedings of the 22nd Annual Conference of the Society for Classification*, pages 252–259, 1999.

[RY90] K. R. Rao et P. Yip. *Discrete Cosine Transform: Algorithms, Advantages, Applications*. Academic Press, Inc, 1990. 490 pages.

[Sam94] F. Samaria. « *Face recognition using Hidden Markov Models* ». Ph.D. thesis, Engineering Department, Cambridge University, 1994.

[Sap90] G. Saporta. *Probabilités, Analyse de données et Statistique*. Editions Technip, 1990. 493 pages.

[SBB03] T. Sim, S. Baker et M. Bsat. « The CMU Pose, Illumination, and Expression Database ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003.

[SBV95] B. Schölkopf, C. Burges et V. Vapnik. « Extracting Support Data for a Given Task ». Dans *Proceedings of the 1st International Conference on Knowledge Discovery & Data Mining*, pages 252–257, 1995.

[SD96] M. Skurichina et R.P.W. Duin. « Stabilizing Classifiers for Very Small Sample Size ». Dans *Proceedings of International Conference on Pattern Recognition (ICPR'96)*, pages 891–896, 1996.

[SD99] M. Skurichina et R.P.W. Duin. « Regularization of Linear Classifiers by Adding Redundant Features ». *Pattern Analysis and Applications*, 2:44–52, 1999.

[SD02] M. Skurichina et R.P.W. Duin. « Bagging, Boosting and the Random Subspace Method for Linear Classifiers ». *Pattern Analysis and Applications*, 5(2):121–135, 2002.

[SFD02] G. Shakhnarovich, J.W. Fisher et T. Darrell. « Face Recognition from Long-Term Observations ». Dans *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 851–868, 2002.

[SH94] F. Samaria et A. Harter. « Parameterisation of a Stochastic Model for Human Face Identification ». Dans *Proceedings of the IEEE Workshop on Applications of Computer Vision*, pages 138–142, 1994.

[SK87] L. Sirovich et M. Kirby. « Low Dimensional Procedure for the Characterization of Human Faces ». *Journal of the Optical Society of America A*, 4(3):519–524, 1987.

[SK00] H. Schneiderman et T. Kanade. « A Statistical Model for 3D Object Detection Applied to Faces and Cars ». Dans *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 746–751, 2000.

[SM04] G. Shakhnarovich et B. Moghaddam. « Face Recognition in Subspaces » . Dans S.Z. Li et A.K. Jain, éditeurs, *Handbook of Face Recognition*, pages 283–297. Springer-Verlag, Reidel, Dordrecht, 2004.

[SP98] K. Sung et T. Poggio. « Example-Based Learning for View-Based Human Face Detection ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):39–51, 1998.

[SS01] B. Schölkopf et A.J. Smola. *Learning with Kernels*. MIT Press, Cambridge, MA., 2001. 626 pages.

[SSM99] Schölkopf B., Smola A.J. et Müller K.R.. « Kernel Principal Component Analysis » . Dans B. Schölkopf, C. Burges et A. Smola, éditeurs, *Advances in Kernel Methods - Support Vector Learning*, pages 327–352. MIT Press, 1999.

[Sto84] T.J. Stonham. « Practical Face Recognition and Verification with WISARD » . Dans H.D. Ellis, M.A. Jeeves, F. Newcombe et A. Young, éditeurs, *Aspects of Face Processing*, pages 426–441. Dordrecht: Nijhoff, 1984.

[SW96] D.L. Swets et J.J. Weng. « Using Discriminant Eigenfeatures for Image Retrieval ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):831–836, 1996.

[Tak98] B. Takács. « Comparing Face Images Using the Modified Hausdorff Distance ». *Pattern Recognition*, 31:1873–1881, 1998.

[TB97] M.E. Tipping et C.M. Bishop. « Probabilistic Principal Component Analysis ». Technical Report NCRG-97-010, Neural Computing Research Group, Aston University, 1997. 13 pages.

[TB98] N.F. Troje et H.H Bülthoff. « How is Bilateral Symmetry of Human Faces Used for Recognition of Novel Views ». *Vision Research*, 38:79–89, 1998.

[TB99a] M.E. Tipping et C.M. Bishop. « Mixtures of Probabilistic Principal Component Analysers ». *Neural Computation*, 11(2):443–482, 1999.

[TB99b] M.E. Tipping et C.M. Bishop. « Probabilistic Principal Component Analysis ». *Journal of the Royal Statistical Society, Series B*, 21(3):611–622, 1999.

[TBGL86] Q. Tian, M. Barbero, Z. Gu et S. Lee. « Image Classification by the Foley-Sammon Transform ». *Optical Engineering*, 25:834–840, 1986.

[TC02] D.S. Turaga et T. Chen. « Face Recognition using Mixtures of Principal Components ». Dans *Proceedings of the International Conference on Image Processing (ICIP)*, volume 2, pages 101–104, 2002.

[TF96] B.G. Tabachnick et L.S. Fidell. *Using Multivariate Statistics*. Harper Collins College Publishers, New York, N.Y., 1996. 880 pages.

[TFV98] C.E. Thomaz, R.Q. Feitosa et A. Veiga. « Design of Radial Basis Function Network as Classifier in Face Recognition Using Eigenfaces ». Dans *5th Brazilian Symposium on Neural Networks (SBRN '98)*, pages 118–123, 1998.

[TG74] J.T. Tou et R.C. Gonzalez. *Pattern Recognition*. Addison-Wesley, 1974. 377 pages.

[TH96] N.F Troje et Bülthoff H.H.. « Face Recognition Under Varying pose: the Role of Texture and Shape ». *Vision Research*, 36:1761–1771, 1996.

[TKC02] Y.L. Tian, T. Kanade et J.F. Cohn. « Recognizing Action Units for Facial Expression Analysis » . Dans *Multimodal Interface for Human-Machine Communication*, pages 32–66. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2002.

[TKP01] A. Tefas, C. Kotropoulos et I. Pitas. « Using Support Vector Machines to Enhance the Performance of Elastic Graph Matching for Frontal Face Authentication ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:735–746, 2001.

[TLV00] L. Torres, L. Lorente et J. Vila. « Automatic Face Recognition of Video

Sequences Using Self-Eigenfaces ». Dans *International Symposium on Image/video Communication over Fixed and Mobile Networks*, 2000. 5 pages.

[TP91] M.A. Turk et A.D. Pentland. « Eigenfaces for Recognition ». *Journal of Cognitive Neuroscience*, 3:71–86, 1991.

[TRL99] L. Torres, J.Y. Reutter et L. Lorente. « The Importance of Color Information in Face Recognition ». Dans *Proceedings of the IEEE International Conference on Image Processing*, volume 3, pages 627–631, 1999.

[TSYQ05] E.K. Tang, P.N. Suganthan, X. Yao et A.K. Qin. « Linear Dimensionality Reduction using Relevance Weighted LDA ». *Pattern Recognition*, 38(4):485–493, Avril 2005.

[Vap95] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995. 314 pages.

[Vap98] V. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, Inc., New York, 1998. 768 pages.

[VGJ04] M. Visani, C. Garcia et J.M. Jolion. « Two-Dimensional-Oriented Linear Discriminant Analysis for Face Recognition ». Dans *Proceedings of the International Conference on Computer Vision and Graphics (ICCVG 2004)*, pages 1008–1017, 2004.

[VGJ05a] M. Visani, C. Garcia et J.M. Jolion. « Bilinear Discriminant Analysis for Face Recognition ». Dans *Proceedings of the 3rd International Conference on Advances in Pattern Recognition (ICAPR 2005)*, volume 2, pages 247–256, Août 2005.

[VGJ05b] M. Visani, C. Garcia et J.M. Jolion. « Face Recognition using Modular Bilinear Discriminant Analysis ». Dans *Proceedings of the 8th International Conference on VISual Information Systems (VIS'05)*, pages 24–34, Juillet 2005.

[VGJ05c] M. Visani, C. Garcia et J.M. Jolion. « Normalized Radial Basis Function Networks and Bilinear Discriminant Analysis for Face Recognition ». Dans *Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS 2005)*, pages 342–347, Septembre 2005.

[VGJ05d] M. Visani, C. Garcia et J.M. Jolion. « Une nouvelle méthode de représentation des visages pour leur reconnaissance : l'Analyse Discriminante Bilinéaire ». Dans *Actes de la Conférence COmpression et REprésentation des Signaux Audiovisuels (CORESA 2005)*, pages 103–108, Novembre 2005.

[VGL04] M. Visani, C. Garcia et C. Laurent. « Comparing Robustness of Two-Dimensional PCA and Eigenfaces for Face Recognition ». Dans *Proceedings of the International Conference on Image Analysis and Recognition (ICIAR'04)*, volume 2, pages 717–724, 2004.

[VJ01] P. Viola et M. Jones. « Rapid Object Detection using a Boosted Cascade of Simple Features ». Dans *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 511–518, 2001.

[VT02] M.A.O. Vasilescu et D. Terzopoulos. « Multilinear Subspace Analysis for Image Ensembles ». Dans *Proceedings of the International Conference on Pattern Recognition (ICPR 2002)*, volume 2, pages 511–514, 2002.

[WFKvdM97] L. Wiskott, J.M. Fellous, N. Krüger et C. von der Malsburg. « Face Recognition by Elastic Bunch Graph Matching ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.

[WJHT04] Y. Wang, Y. Jia, C. Hu et M. Turk. « Face Recognition Based on Kernel Radial Basis Function Networks ». Dans *Asian Conference on Computer Vision*, 2004. 6 pages.

[WS03] L. Wolf et A. Shashua. « Learning Over Sets using Kernel Principal Angles ». *Journal of Machine Learning Research*, 4:913–931, 2003.

[WT04a] X. Wang et X. Tang. « Dual-Space Linear Discriminant Analysis for Face Recognition ». Dans *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 564–569, 2004.

[WT04b] X. Wang et X. Tang. « Random sampling LDA for face recognition ». Dans *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 259–265, 2004.

[WWZF05] L. Wang, X. Wang, X. Zhang et J. Feng. « The Equivalence of Two-Dimensional PCA to Line-Based PCA ». *Pattern Recognition Letters*, 26:57–60, Janvier 2005.

[XKY94] L. Xu, A. Krzyżak et A. Yuille. « On Radial Basis Function Nets and Kernel Regression: Approximation Ability, Convergence Rate and Receptive Field Size ». *Neural Networks*, 7:609–628, 1994.

[XSA05] H. Xiong, M.N.S. Swamy et M.O. Ahmad. « Two-dimensional FLD for Face Recognition ». *Pattern Recognition*, 38(7):1121–1124, Janvier 2005.

[Yam00] W.S. Yambor. « Analysis of PCA-based and Fisher Discriminant-Based Image Recognition Algorithms ». Technical report cs-00-103, Computer Science Department, Colorado State University, 2000. 76 pages.

[Yan02] M.H. Yang. « Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods ». Dans *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 215–220, 2002.

[YDB00] W.S. Yambor, B.A. Draper et J.R. Beveridge. « Analyzing PCA-based Face Recognition Algorithms: Eigenvector Selection and Distance Measures ». Dans *2nd Workshop on Empirical Evaluation in Computer Vision*, 2000. 14 pages.

[YFM98] O. Yamaguchi, K. Fukui et K.I. Maeda. « Face Recognition using Temporal Image Sequence ». Dans *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 318–323, 1998.

[YJPP04] J. Ye, R. Janardan, C.H. Park et H. Park. « An Optimization Criterion for Generalized Discriminant Analysis on Undersampled Problems ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:982–994, 2004.

[YKA01] M. Yang, D. Kriegman et N. Ahuja. « Face Detection Using Multimodal Density Models ». *Computer Vision and Image Understanding*, 84:264–284, 2001.

[YKA02] M.H. Yang, D.J. Kriegman et N. Ahuja. « Detecting Faces in Images: A Survey ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.

[YY01] H. Yu et J. Yang. « A Direct LDA Algorithm for High Dimensional Data - with Application to face Recognition ». *Pattern Recognition*, 34:2067–2070, 2001.

[YZFY04] J. Yang, D. Zhang, A.F. Frangi et J.Y. Yang. « Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):131–137, 2004.

[YZY03] J. Yang, D. Zhang et J. Yang. « A Generalized KL expansion Method Which can Deal with Small Sample Size and High-Dimensional Problems ». *Pattern Analysis and Applications*, 6:47–54, 2003.

[ZCK98] W. Zhao, R. Chellappa et A. Krishnaswamy. « Discriminant Analysis of Principal Components for Face Recognition ». Dans *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 336–341, 1998.

[ZCPR03] W. Zhao, R. Chellappa, P.J. Phillips et A. Rosenfeld. « Face Recognition: a Literature Survey ». *ACM Computing Surveys*, 35(4):399–458, 2003.

[ZH01] X.S. Zhou et T.S. Huang. « Small Sample Learning During Multimedia Retrieval using BiasMap ». Dans *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 11–17, 2001.

[Zha99] W. Zhao. « *Robust Image Based 3D Face Recognition* ». PhD thesis, University of Maryland, 1999.

[Zho05] S.K. Zhou. « Matrix-Based Kernel Subspace Methods ». 2005. Disponible sur http://www.cfar.umd.edu/ shaohua/papers/zhou05tpami_mtx.pdf, 15 pages.

[ZSL05] D. Zhang, S. Songcan et J. Liu. « Representing Image Matrices: Eigenimages vs. Eigenvectors ». Dans *Proceedings of the 2nd International Symposium on Neural Networks (ISNN'05)*, pages 659–664, Juin 2005.

Références de l'auteur

Conférences internationales avec comité de lecture et actes :

M. Visani, C. Garcia, J.M. Jolion, « Normalized Radial Basis Function Networks and Bilinear Discriminant Analysis for Face Recognition ». Dans *Proceedings of the IEEE International Conference on Advanced Video and Signal based Surveillance* (AVSS 2005), IEEE Computer Society, pages 342–347, Côme, Italie, Septembre 2005.

M. Visani, C. Garcia, J.M. Jolion, « Face Recognition using Modular Bilinear Discriminant Analysis ». Dans *Proceedings of the International Conference on Visual Information Systems* (VIS 2005). Springer LNCS 3736, S. Bres, R. Laurini (eds), pages 24–34, Amsterdam, Pays-Bas, Juillet 2005.

M. Visani, C. Garcia, J.M. Jolion, « Bilinear Discriminant Analysis for Face Recognition ». Dans *Proceedings of the International Conference on Computer Vision and Graphics* (ICAPR 2005). Springer LNCS 3687, S. Singh, M. Singh, C. Apte, P. Perner (eds), volume 2, pages 247–256, Bath, Angleterre, Août 2005.

M. Visani, C. Garcia, J.M. Jolion, « Two-Dimensional-Oriented Linear Discriminant Analysis for Face Recognition ». Dans *Proceedings of the International Conference on Computer Vision and Graphics* (ICCVG 04). Série *Computational Imaging and Vision*, pages 1008–1017, Varsovie, Pologne, Septembre 2004.

M. Visani, C. Garcia, C. Laurent, « Comparing Robustness of Two-Dimensional PCA and Eigenfaces for Face Recognition ». Dans *Proceedings of the International Conference on Image Analysis and Recognition* (ICIAR 04). Springer LNCS 3211, A. Campilho, M. Kamel (eds), volume 2, pages 717–724, Porto, Portugal, Octobre 2004.

C. Laurent, N. Laurent, M. Visani, « Color Image Retrieval Based on Wavelet Salient Features Detection ». Dans *Proceedings of the International Workshop on Content-Based Multimedia Indexing* (CBMI 03), pages 327–334, Rennes, France, Septembre 2003.

Conférences nationales avec comité de lecture et actes :

M. Visani, C. Garcia, J.M. Jolion, « Une Nouvelle Méthode de Représentation des Visages pour leur Reconnaissance : l'Analyse Discriminante Bilinéaire ». Dans *Actes de la Conférence Compression et REprésentation des Signaux Audiovisuels* (CORESA 2005) pages 103–108, Rennes, France, Novembre 2005.

Brevet :

M. Visani, C. Garcia, C. Laurent, « Analyse Discriminante Linéaire Bi-dimensionnelle pour la Reconnaissance de Visages ». Numéro de dépôt : 04/01395, voie internationale PCT, France Télécom, 04 Juin 2004.

FOLIO ADMINISTRATIF

THESE SOUTENUE DEVANT L'INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE LYON

NOM : **VISANI** DATE de SOUTENANCE : 25/11/2005

Prénoms : **Muriel**

NATURE : Doctorat Numéro d'ordre : 2005-ISAL-0094

Ecole doctorale : Ecole Doctorale Informatique et Information pour la Société (EDIIS)

Spécialité : **Informatique**

Laboratoire (s) de recherche : Laboratoire d'InfoRmatique en Image et Systèmes d'information **(LIRIS)**

Cote B.I.U. - Lyon : T 50/210/19 / et bis CLASSE :

TITRE : Vers de nouvelles approches discriminantes pour la reconnaissance automatique de visages

RESUME : Les travaux effectués dans le cadre de cette thèse portent sur l'identification automatique de visages dans des images numériques. L'objectif est d'assigner à des visages-requêtes une identité parmi celles d'un ensemble de personnes connues. Pour cela, on cherche à extraire, pour chaque visage, un ensemble de descripteurs appelé signature qui lui soit spécifique, puis à définir un schéma de classification des signatures adapté à l'application visée. De nombreuses méthodes ont été proposées dans la littérature. Parmi les plus efficaces, on compte les techniques de projection statistique, dont le but est de fournir, par le biais d'une analyse multidimensionnelle des données, un espace de représentation plus adapté à la classification que l'espace initial des données.
Ce travail reprend ce principe et propose de nouvelles techniques d'extraction de signatures basées sur l'Analyse Discriminante Linéaire qui, contrairement à la plupart des approches existantes, prennent en compte la structure bidimensionnelle des images de visages. Les méthodes proposées permettent de pallier les principaux désavantages des techniques usuelles. Elles contournent le problème de la singularité sans nécessiter l'ajout d'aucun paramètre et leur construction est moins coûteuse et instable. Un schéma original de classification des signatures ainsi obtenues, en monde fermé ou ouvert, est également introduit. Les techniques proposées sont évaluées et comparées aux approches usuelles selon des protocoles expérimentaux rigoureux. Les résultats ainsi obtenus montrent leurs très bonnes performances, et notamment une robustesse accrue vis-à-vis de changements de pose ou d'expression faciale et d'occultations partielles.

MOTS-CLES : Analyse d'images, reconnaissance de formes, classification automatique, analyse discriminante.

TITLE : Towards new discriminant approaches for automatic face recognition

ABSTRACT : This thesis is concerned with automatic identification of human faces from images. The objective is to assign an identity to any query face among a set of known people. The underlying systems aim at extracting a specific set of descriptors called signature for every face and at designing a signature classification scheme adapted to the context. Statistical projection-based methods are among the most effective techniques that have been proposed. They use multidimensional data analysis methods to define a new projection space in which classification can be more effectively performed.
In this thesis, this principle is used to define new feature extraction techniques using Linear Discriminant Analysis. Unlike traditional approaches, the 2D structure of the images is taken into account when defining the projection space. The proposed methods allow to overcome the main drawbacks of the state-of-the-art methods in terms of numerical cost and instability when building the system. Moreover, they get rid of the singularity problem and do not involve any additional parameter. An original classification scheme is also proposed for classifying signatures in closed-world and open-world environments. The proposed techniques are evaluated and compared to usual methods using various international face databases. These experiments show the effectiveness of the proposed approaches. In particular, they highlight an increased robustness to head pose changes, facial expression variations and partial occlusions.

KEYWORDS : Image Analysis, Pattern Recognition, Automatic Classification, Discriminant Analysis.

Président du jury : Jean-Michel POGGI

Directeurs de thèse : Jean-Michel JOLION et Christophe GARCIA

Composition du jury : Jean-Michel POGGI (Président), Stéphane LALLICH (Rapporteur), Jack-Gérard POSTAIRE (Rapporteur), Jean-Michel JOLION (Directeur), Christophe GARCIA (Directeur), Jean-Luc DUGELAY (Examinateur).