



HAL
open science

Nouvelle approche d'identification dans les bases de données biométriques basée sur une classification non supervisée

Anis Chaari

► **To cite this version:**

Anis Chaari. Nouvelle approche d'identification dans les bases de données biométriques basée sur une classification non supervisée. Modélisation et simulation. Université d'Evry-Val d'Essonne, 2009. Français. NNT: . tel-00549395

HAL Id: tel-00549395

<https://theses.hal.science/tel-00549395>

Submitted on 21 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université d'Evry Val d'Essonne
UFR Sciences et technologies



Université de la Manouba
Ecole Nationale des Sciences de l'Informatique



THESE

Présenté par

Anis CHAARI

Pour obtenir le diplôme du doctorat

Spécialités Sciences pour l'ingénieur et Informatique

Sujet

Nouvelle approche d'identification dans les bases de données biométriques basée sur une classification non supervisée

Préparé au sein des laboratoires IBISC et RIADI



Soutenue le mardi 6 Octobre 2009 au laboratoire IBISC

Composition du jury

Philippe Bolon	Professeur	Université de Savoie	Président
Bernadette Dorizzi	Professeur	Télécom et Management Sud Paris	Rapporteur
Jean Marc Ogier	Professeur	Université de la Rochelle	Rapporteur
Kamel Hamrouni	MCF (HDR)	Université Tunis El Manar	Rapporteur
Patrick Bonnin	Professeur	Université de Versailles	Examineur
Sylvie Lelandais Bonadè	Professeur	Université d'Evry	Directeur de thèse
Mohamed Ben Ahmed	Professeur Emérite	Université de la Manouba	Directeur de thèse

Remerciements

C'est avec grand plaisir que je réserve cette page, en signe de gratitude et de reconnaissance à tous ceux qui m'ont aidé à la réalisation de ce travail.

Je remercie, tout d'abord, mes Professeurs Mohamed Ben Ahmed et Sylvie Lelandais-Bonade, pour leurs confiances, leurs disponibilités et leurs soutiens continuels. Qu'ils trouvent ici l'expression de ma gratitude pour leurs précieux conseils et toute l'aide qu'ils m'ont procurée durant l'élaboration de ce travail. Je tiens à les remercier bien vivement pour leurs encadrements, leurs encouragements et aussi pour leurs contributions à l'amélioration judicieuse de la qualité de ce mémoire.

Je remercie vivement les membres de jury d'avoir accepté de juger mon travail de thèse. Je suis très reconnaissant à Bernadette Dorizzi, Jean Marc Ogier et Kamel Hamrouni de s'être penchés avec rigueur et grand intérêt sur ce rapport et de m'avoir fait l'honneur d'en être les rapporteurs. Mes plus sincères remerciements vont aussi à Philippe Bolon et Patrick Bonnin qui ont examiné ce travail de thèse et ont participé au jury.

Ce travail de thèse a été réalisé dans le cadre d'une co-tutelle entre le laboratoire RIADI-GDL de l'Ecole Nationale des Sciences de l'Informatique de Tunis et le Laboratoire IBISC (FRE CNRS 3190) de l'Université d'Evry. Je remercie leurs directeurs ainsi que tous les chercheurs et l'ensemble du personnel de ces laboratoires pour m'avoir fait profiter de leur expérience et de leur savoir.

Je remercie également Mohamed Saaidia de l'université d'Annaba avec qui j'ai collaboré pour effectuer une contribution non négligeable sur la localisation de visages.

Je n'oublie pas non plus de remercier toutes les personnes que j'ai pu rencontrer et avec lesquelles j'ai pu échanger, et en particulier les collaborateurs du projet IV². Je remercie également l'ensemble du personnel du département QLIO de l'IUT d'Evry pour leurs accueils chaleureux et les conseils pédagogiques dont ils m'ont fait part.

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

Je remercie, enfin, toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce travail.

Résumé

Les travaux effectués dans le cadre de cette thèse portent sur l'identification automatique de visages dans des bases de données d'images numériques. L'objectif est de simplifier le déroulement de l'identification biométrique qui consiste à la recherche de l'identité requête parmi l'ensemble des individus de la base de données préenregistrée, appelée aussi galerie. En effet, le schéma d'identification classique est complexe et très coûteux notamment dans le cas de grandes bases de données biométriques. Le processus original que nous proposons ici a pour objectif de réduire la complexité et d'améliorer les performances en terme de temps de calcul et de taux d'identification.

Dans ce cadre biométrique, nous avons proposé une étape de classification non supervisée ou clustering des visages pour partitionner la base d'images en plusieurs sous ensembles cohérents et bien discriminés. Pour ce faire, l'algorithme de clustering vise à extraire, pour chaque visage, un ensemble de descripteurs, appelé signature, qui lui soit spécifique. Trois techniques de représentation faciales ont été développées dans le but d'extraire des informations différentes et complémentaires décrivant le visage humain : deux méthodes factorielles d'analyse multidimensionnelle et de projection des données (méthodes dites « *Eigenfaces* » et « *Fisherfaces* ») ainsi qu'une méthode d'extraction des moments géométriques de Zernike. Sur la base des différentes classes de signatures extraites pour chaque visage plusieurs méthodes de clustering sont mises en compétition afin de dégager la classification optimale qui conduit à une meilleure réduction de la galerie. Les méthodes retenues sont, soit de type « centres mobiles » telles que les K-moyennes et les centres mobiles de Forgy, ou de type « agglomérative » telle que la méthode de BIRCH. Sur la base de la dépendance des partitions générées, ces différentes stratégies classificatoires sont ensuite combinées suivant une architecture parallèle de manière à optimiser la réduction de l'espace de recherche à un sous ensemble de la base de données. Les clusters retenus in fine étant ceux pour lesquels la probabilité d'appartenance de l'identité requête est quasi certaine.

Mots clés: Analyses d'images, signature biométrique, classification non supervisée (clustering), base de données d'images, identification de visages.

Abstract

The work done in the framework of this thesis deal with the automatic faces identification in databases of digital images. The goal is to simplify biometric identification process that is seeking the query identity among all identities enrolled in the database, also called gallery. Indeed, the classical identification scheme is complex and requires large computational time especially in the case of large biometric databases. The original process that we propose here aims to reduce the complexity and to improve the computing time and the identification rate performances.

In this biometric context, we proposed an unsupervised classification or clustering of facial images in order to partition the enrolled database into several coherent and well discriminated subsets. In fact, the clustering algorithm aims to extract, for each face, a specific set of descriptors, called signature. Three facial representation techniques have been developed in order to extract different and complementary information which describe the human face: two factorial methods of multidimensional analysis and data projection (namely called "*Eigenfaces*" and "*Fisherfaces*") and a method of extracting geometric Zernike moments. On the basis of the different signatures obtained for each face, several clustering methods are used in competing way in order to achieve the optimal classification which leads to a greater reduction of the gallery. We used either "mobile centers" methods type such as the K-means algorithm of MacQueen and that of Forgy, and the "agglomerative" method of BIRCH. Based on the dependency of the generated partitions, these different classifying strategies are then combined using a parallel architecture in order to maximize the reduction of the search space to the smallest subset of the database. The retained clusters in fine are those which contain the query identity with an almost certain probability.

Key words: Image analysis, biometric signature, unsupervised classification (clustering), images databases, facial identification.

Table des matières

Introduction Générale.....	1
Contexte et motivations.....	1
Contribution et plan du manuscrit.....	2
Chapitre I Problématique de l'identification dans les bases de données biométriques....	5
1.1 Introduction	5
1.2 La reconnaissance biométrique	5
1.3 Les caractéristiques biométriques	6
1.4 Conception d'un système biométrique.....	8
1.4.1 Mode de reconnaissance.....	10
1.4.2 Applications et enjeux.....	11
1.4.3 Applications visées.....	13
1.5 Mesure des performances des systèmes de reconnaissance biométriques	13
1.5.1 Erreur des systèmes d'authentification	14
1.5.2 Evaluation des systèmes d'identification	17
1.5.3 Les intervalles de confiance	20
1.6 Position du problème d'identification	20
1.6.1 Travaux d'évaluation large échelle	21
1.6.2 Limitations et solution proposée	23
1.7 Etat de l'art de systèmes de classification des signatures biométriques	25
1.7.1 Systèmes de classification des empreintes digitales	25
1.7.2 Classification des images faciales suivant le genre et/ou la race ethnique	26
1.7.3 Classification non supervisée pour la reconnaissance biométrique	27
1.8 Conclusion.....	27
Chapitre II Détection et localisation de visages - Vers la reconnaissance automatique des visages	29
2.1 Introduction	29
2.2 Chaîne de traitement préalable à la reconnaissance automatique	29
2.3 La détection et la reconnaissance faciale - un processus d'apprentissage	30
2.4 Détection et localisation de visage - Etat de l'art.....	31
2.4.1 Les approches locales.....	32
2.4.2 Les approches globales.....	32
2.5 Notre approche de localisation.....	32
2.5.1 Les moments de Zernike	34
2.5.2 Introduction aux « <i>Eigenfaces</i> ».....	37
2.5.3 Méthode de localisation par réseau de neurones.....	39
2.5.4 Qualité de la Localisation.....	42
2.6 Résultats expérimentaux	44
2.6.1 Etude comparative de la caractérisation Moments de Zernike - <i>Eigenfaces</i>	44
2.7 Normalisation des visages.....	46
2.8 Conclusion.....	48
Chapitre III Reconnaissance faciale - Etat de l'art des techniques de reconnaissance ..	49
3.1 Introduction	49
3.2 Positionnement.....	49
3.3 La reconnaissance humaine de visages	50
3.3.1 La reconnaissance humaine de visages : un processus complexe.....	51

3.3.2	Processus local et global	52
3.3.3	Impact des différents types de variabilités	52
3.4	Difficultés inhérentes à la reconnaissance automatique.....	53
3.4.1	Les variations de la pose	53
3.4.2	Les changements d'éclairage	53
3.4.3	Les expressions faciales	54
3.4.4	Les occultations partielles	54
3.4.5	Le vieillissement et le changement d'aspect	55
3.5	Les méthodes de reconnaissance automatique de visages	55
3.5.1	Les approches globales.....	56
3.5.2	Les approches d'identification basées sur les caractéristiques locales et hybrides	62
3.6	Conclusion.....	67
Chapitre IV Bases de données utilisées et corpus d'évaluation.		69
4.1	Introduction	69
4.2	La base IV ²	70
4.3	Les évaluations menées dans le cadre du projet IV ²	71
4.3.1	La phase de développement	71
4.3.2	Conception et mise en œuvre des évaluations.....	71
4.3.3	Conclusion sur les évaluations	77
4.4	Les bases de visages utilisées.....	77
4.4.1	La base XM2VTS [Mes99].....	78
4.4.2	La base FERET	78
4.4.3	La base AR [MB98]	79
4.4.4	La base CVL [CVL.si]	80
4.4.5	La base ORL.....	80
4.4.6	Notre corpus d'évaluation	80
4.5	Conclusion.....	81
Chapitre V Nouvelle approche d'identification dans les bases de données biométriques - Application à la reconnaissance faciale 2D		83
5.1	Introduction	83
5.2	Les différentes étapes de notre approche d'identification.....	83
5.3	Les données biométriques	85
5.3.1	Les métadonnées	85
5.3.2	Organisation des données.....	85
5.3.3	Stratégies expérimentales	86
5.3.4	Résultats d'identification classique.....	88
5.4	Représentation des visages 2D.....	90
5.4.1	Les méthodes factorielles	90
5.4.2	Les moments de Zernike	93
5.4.3	Critères d'extraction des attributs	94
5.4.4	Extraction des vecteurs attributs	95
5.5	Classification et partitionnement des vecteurs caractéristiques	101
5.5.1	Schéma général de la classification automatique.....	101
5.5.2	Structures Classificatoires	102
5.5.3	Mise en œuvre de la classification des visages	104
5.5.4	Résultats de classification	106
5.6	Combinaison Classification - Identification.....	114

5.7	Conclusion.....	118
Chapitre VI Généralisation des données faciales et des méthodes de classification pour notre approche d'identification..... 119		
6.1	Introduction	119
6.2	Approche conceptuelle d'identification biométrique.....	119
6.3	Corpus de données	122
6.3.1	Choix de l'espace d'apprentissage	122
6.4	Extraction des vecteurs attributs	123
6.4.1	Choix des attributs <i>Eigenfaces</i>	124
6.4.2	Choix des attributs <i>Fisherfaces</i>	125
6.4.3	Choix des moments de Zernike	126
6.4.4	Synthèse des vecteurs attributs extraits	126
6.5	Partitionnement de la base de données.....	126
6.5.1	Les centres mobiles de Forgy.....	127
6.5.2	La méthode BIRCH.....	128
6.5.3	Performances des différentes méthodes de classification	131
6.6	Etude d'indépendance des différents classifieurs.....	137
6.6.1	Indépendance des attributs	138
6.6.2	Indépendance des méthodes de classification et des mesures de similarité...	139
6.6.3	Choix de la distance de similarité	140
6.7	Agrégation et combinaison de classifieurs.....	140
6.7.1	Les architectures de combinaison	141
6.7.2	Les différents niveaux de fusion pour un système multimodal.....	141
6.7.3	Notre stratégie de fusion	142
6.7.4	Résultats de la fusion.....	143
6.8	Identification finale	144
6.8.1	Précision de l'identification	145
6.8.2	Temps de réponse.....	148
6.9	Conclusion.....	150
Conclusion Générale		153
Références		157
ANNEXES.....		171
Annexe A Développement d'une base de données biométriques : le projet IV².....		173
A.1	Conception et réalisation de la cabine d'acquisition.....	173
A.2	Les protocoles d'acquisition.....	177
Annexe B Centre, rayon et diamètre d'une classe.....		181
Annexe C Classes d'individus		183
Annexe D Algorithme d'insertion d'une entrée dans un CF-tree.....		185
Annexe E Implémentation de BIRCH.....		187

Table des Figures

Figure 1.1 : Les principales caractéristiques biométriques pouvant servir comme moyen d'identification : a) forme de l'oreille, b) visage 2D, c) visage 3D, d) visage infrarouge, e) iris, f) rétine, g) empreinte de la main, h) thermogramme de la main, i) forme de la main, j) empreinte digitale, k) voix, l) signature et m) réseau veineux de la main [MMJP05].	7
Figure 1.2 : Schéma de fonctionnement d'un système biométrique. Diagrammes des processus d'enrôlement, de vérification et d'identification.	9
Figure 1.3 : La distribution des scores des personnes légitimes et des imposteurs et les taux d'erreurs, FAR et FRR , pour un seuil donné t .	15
Figure 1.4 : Courbe DET.	16
Figure 1.5 : Courbe ROC.	16
Figure 1.6 : Exemple des courbes $FAR(t)$ et $FRR(t)$, où les points correspondant aux EER , $ZeroFRR$ et $ZeroFAR$ sont illustrés.	16
Figure 1.7 : Exemple de courbe de scores cumulés (CMS).	17
Figure 1.8: Courbes de scores cumulés, communiqués par «FRVT 2002 Evaluation Report» [PGM ⁺ 03], pour une très grande base de données de 37437 personnes.	22
Figure 1.9 : Taux d'identification en fonction de la taille de la galerie, communiqués par «FRVT 2002 Evaluation Report» [PGM ⁺ 03].	23
Figure 1.10 : Les cinq classes des empreintes digitales.	26
Figure 2.1 : Prétraitements des images en amont de la reconnaissance. Les différentes étapes du prétraitement peuvent également servir pour d'autres applications (encadrés en pointillés).	30
Figure 2.2 : Schéma de principe de notre méthode de localisation du visage.	34
Figure 2.3 : Façon de recalculer de la fonction $h(x,y)$ (rectangle gris) dans le cercle unité : a) cas de $c = -1$ et $d = 1$; b) cas de $c = -1/\sqrt{2}$ et $d = 1/\sqrt{2}$.	36
Figure 2.4 : a) Exemple d'images utilisées dans l'ensemble d'apprentissage, b) Visage moyen.	37
Figure 2.5 : Les six premiers <i>Eigenfaces</i> calculés à partir de l'ensemble d'images de la figure 2.4.	38
Figure 2.6 : Exemple de localisation par la méthode des <i>Eigenfaces</i> .	39
Figure 2.7. Réseau de neurones multicouches (Perceptron Multicouche PMC ; en anglais Muti-layer Perceptron MLP).	40
Figure 2.8 : Phase d'apprentissage de la procédure de localisation faciale par réseau de neurones.	41
Figure 2.9 : Phase de test de la procédure de localisation faciale par réseau de neurones.	41
Figure 2.10 : Images originales et les trois zones caractéristiques associées à chacune d'elles.	43
Figure 2.11: Exemples sur la complémentarité des deux mesures Gdr et Qdr .	43
Figure 2.12 : Courbes des taux de localisation Gdr (en haut) et Qdr (en bas) pour chaque image test. Ces taux sont obtenus par un réseau de neurones entraîné par les moments de Zernike (à gauche) et les <i>Eigenfaces</i> (à droite).	45
Figure 2.13 : Résultats de localisation et valeurs de Gdr et Qdr obtenus pour des images de la base XM2VTS. En haut, images d'origine. Au milieu, localisation par l'algorithme utilisant les moments de Zernike. En bas, localisation par les <i>Eigenfaces</i> .	45
Figure 2.14 : Nombres de pixels retenus pour le découpage des images de visages	47
Figure 2.15 : Le processus de normalisation des images compte quatre étapes : la rotation, le redimensionnement, le découpage, et l'égalisation d'histogramme.	47
Figure 3.1 : Comparaison des <i>Eigenfaces</i> dual et des <i>Eigenfaces</i> standard: (a) interpersonnel, (b) extrapersonnel, (c) standard (Moghaddam et Pentland [MP97]).	58
Figures 3.2: Exemple de cinq premiers <i>Fisherfaces</i> , associés aux plus grandes valeurs propres. Extrait de [Vis05].	59
Figure 3.3: Extrait de [DBBB03]. Les six premiers axes de l'espace de visage obtenus (a) grâce à l'architecture I et (b) avec l'architecture II.	60

Figure 3.4. L'image du visage donne les positions des centres des différentes grilles locales. Les images (a) à (e) montrent les noyaux locaux associés à ces grilles. Extrait de Penev et Atick [PA96].	65
Figure 3.5. Représentation élastique de graphe. Extrait de Wiskott et al. [Wis97].	66
Figure 3.6. Un Face Bunch Graph (FBG) qui vise à représenter l'ensemble des états possibles, pour chaque nœud. Extrait de Wiskott et al. [Wis97].	67
Figure 4.1 : Histogrammes des distances intra-classe & inter-classe.	74
Figure 4.2 : Courbes DET réalisées avec une échelle logarithmique et linéaire.	75
Figure 4.3 : Exemple d'une courbe DET regroupant les différents résultats des participants à un même test (ici visages 2D et Test03).	76
Figure 4.4: Exemples d'images de la base XM2VTS.	78
Figure 4.5 : Extrait de la base Color FERET Les images sont transformées en niveau de gris.	79
Figure 4.6 : Extrait de la base AR. Ensembles des vues collectées pour un individu de la base.	79
Figure 4.7 : Images faciales extraites de la base CVL.	80
Figure 4.8 : 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'éclairage.	80
Figure 5.1 : les différentes étapes de notre approche d'identification biométrique. Les étapes en gris constituent les originalités proposées par rapport aux systèmes d'identification classiques.	84
Figure 5.2 : Organisation des données pour la reconnaissance faciale 2D.	86
Figure 5.3 : Courbes de Scores Cumulés des <i>Eigenfaces</i> entraînés différemment avec des ensembles de 600, 800, 1200 et 1600 images de la base XM2VTS.	88
Figure 5.4 : Courbes de Scores Cumulés des <i>Eigenfaces</i> avec différents ensembles d'enrôlement.	89
Figure 5.5 : Allure typique des valeurs propres rangées par ordre décroissant.	91
Figure 5.6: illustration du taux d'identification et du rang d'identification totale sur une courbe de scores cumulés.	95
Figure 5.7 : Taux d'identification pour les différentes tailles des vecteurs caractéristiques. Ces différentes tailles correspondent à des inerties allant de 22% à 100% de l'espace de visages.	96
Figure 5.8 : Rang d'identification totale pour les différentes tailles des vecteurs caractéristiques. Ces différentes tailles correspondent à des inerties allant de 22% à 100% de l'espace de visages.	97
Figure 5.9 : Taux d'identification en fonction de l'ordre des moments de Zernike.	99
Figure 5.10 : Rang d'identification totale en fonction de l'ordre des moments de Zernike.	100
Figure 5.11 : Différentes étapes de la classification automatique, extrait de [BC07].	102
Figure 5.12 : Exemples de structures classificatoires.	103
Figure 5.13 : Pseudo code de l'algorithme d'initialisation des centroïdes.	105
Figure 5.14 : Images de visages bruitées progressivement. Le pourcentage du bruitage figurant est calculé comme étant le nombre des pixels bruités aléatoirement sur le nombre total des pixels de l'image.	108
Figure 5.15: Taux d'identification et de classification en fonction du degré de bruitage des images de test.	108
Figure 5.16: Représentation des quatre premières classes suite à une clusterisation des vecteurs <i>Eigenfaces</i> .	109
Figure 5.17 : Courbes de classes cumulées (BCMS) avec une classification en 10 et en 30 classes.	111
Figure 5.18 : Taux de pénétration maximal, moyen et minimal (parmi 400 images faciales de test) des vecteurs caractéristiques <i>Eigenfaces</i> en fonction du nombre de classes.	112
Figure 5.19 : Comparaison des taux de pénétration moyen des <i>Eigenfaces</i> , des <i>Fisherfaces</i> et des Moments de Zernike.	112
Figure 5.20 : Comparaison du temps de calcul de notre approche d'identification et le temps moyen de l'identification classique en fonction du nombre de classes calculées.	113
Figure 5.21 : Comparaison des courbes de scores cumulés de notre approche et celle de l'identification classique. La classification est réalisée avec les <i>Eigenfaces</i> ($g = 44$) alors que l'identification est réalisée avec les <i>Eigenfaces</i> ($g = 190$).	116

Figure 5.22 : Comparaison des courbes de scores cumulés de notre approche et celle de l'identification classique. La classification est réalisée avec les <i>Eigenfaces</i> ($g = 44$) alors que l'identification est réalisée avec les <i>Fisherfaces</i> ($g = 90$).	116
Figure 5.23 : Comparaison des courbes de scores cumulés de notre approche et celle de l'identification classique. La classification est réalisée avec les <i>Eigenfaces</i> ($g = 44$) alors que l'identification est réalisée avec les moments de Zernike (ordre = 30).	117
Figure 6.1 : Approche générique d'identification dans les bases de données biométriques	120
Figure 6.2 : Taux d'identification de la méthode <i>Eigenfaces</i> en fonction de la taille de l'ensemble d'apprentissage. Cet ensemble d'apprentissage est requis pour la construction de l'espace de projection pour les méthodes <i>Eigenfaces</i> et <i>Fisherfaces</i> .	123
Figure 6.3 : Evolution du taux d'identification (IR) et du rang d'identification totale (TMR) des <i>Eigenfaces</i> en fonction de l'inertie globale du corpus de travail.	124
Figure 6.4 : Evolution du taux d'identification (IR) et du rang d'identification totale (TMR) des <i>Fisherfaces</i> en fonction de l'inertie de la dimension sur notre corpus de travail.	125
Figure 6.5 : Evolution du taux d'identification (IR) et du rang d'identification totale (TMR) des Moments de Zernike en fonction de l'inertie de la dimension sur notre corpus de travail.	126
Figure 6.6 : Algorithme BIRCH.	130
Figure 6.7 : Taux de pénétration des <i>Eigenfaces</i> (bleu), des <i>Fisherfaces</i> (vert) et des moments de Zernike (rouge) suite à une classification par les K-moyennes.	132
Figure 6.8 : Taux de pénétration des <i>Eigenfaces</i> (bleu), des <i>Fisherfaces</i> (vert) et des moments de Zernike (rouge) suite à une classification automatique par les centres mobiles de Forgy.	134
Figure 6.9 : Combinaison a) en parallèle, b) en série, c) hiérarchique de N classifieurs	141
Figure 6.10 : Déroulement de l'étape d'identification (En ligne) en aval de la classification non supervisée.	145
Figure 6.11 : Comparaison des courbes de scores cumulés de notre approche et celle de l'identification classique. La classification est réalisée avec BIRCH sur les attributs <i>Fisherfaces</i> ($g = 4$) alors que l'identification est réalisée avec les attributs <i>Eigenfaces</i> ($g = 131$).	146
Figure 6.12 : Comparaison des courbes de scores cumulés de notre approche et celle de l'identification classique. La classification est réalisée avec les K-moyennes sur les attributs <i>Eigenfaces</i> ($g = 16$) alors que l'identification est réalisée avec les <i>Fisherfaces</i> ($g = 67$).	146
Figure 6.13 : Comparaison des courbes de scores cumulés de notre approche et celle de l'identification classique. La classification est réalisée avec les K-moyennes sur les moments de Zernike (ordre = 14) alors que l'identification est réalisée avec les <i>Eigenfaces</i> ($g = 131$).	147
Figure 6.14 : Comparaison des courbes de scores cumulés de notre approche et celle de l'identification classique. La fusion des partitions générées par BIRCH et les K-moyennes implique le sous ensemble d'identités retenu pour l'identification finale. Celle-ci est réalisée avec les vecteurs attributs <i>Eigenfaces</i> ($g = 131$).	148
Figure A.1 : Vue de dessus, et vue de profil de la cabine.	174
Figure A.2 : Structure de la cabine et disposition de l'éclairage.	175
Figure A.3 : « Photomaton » biométrique du projet IV ² .	175
Figure A.4 : Une caméra et un des bancs stéréoscopiques installés dans la cabine.	175
Figure A.5 : Le scanner permettant l'acquisition des visages 3D.	176
Figure A.6 : Le caméscope et la webcam (au milieu du banc stéréoscopique de face) permettant d'enregistrer la modalité visage parlant.	176
Figure A.7 : La caméra infrarouge permettant l'acquisition de l'iris.	177
Figure A.8 : Acquisition d'une face par le scanner 3D	178
Figure A.9 : Images provenant des séquences prises par le caméscope numérique et par la webcam.	179
Figure A.10 : Les deux vues délivrée par les caméras de la base stéréoscopique « face »	180
Figure A.11 : Les deux vues délivrée par les caméras de la base stéréoscopique « profil »	180
Figure A.12 : Photo d'un iris prise par la caméra infrarouge	180

Liste des tableaux

Tableau 1.1 : Avantages et inconvénients des technologies biométriques [Clu04].	8
Tableau 1.2 : Les différentes catégories d'applications de reconnaissances biométriques vis à vis de la tâche et du mode de reconnaissance, la taille de la base de données, le niveau de sécurité, ainsi que le contexte [Vis05].	12
Tableau 1.3 : Performance typique de chaque modalité biométrique évaluée pour des <i>FRR</i> de l'ordre de 1% [JPP ⁺ 04].	24
Tableau 2.1 : Résultats des deux mesures de qualité de localisation de visage obtenus par les moments de Zernike et les <i>Eigenfaces</i> .	46
Tableau 3.1 : Comparaison des pseudos moments de Zernike avec les <i>Eigenfaces</i> et les <i>Fisherfaces</i> . Extrait de [PTN06].	61
Tableau 3.2 : Comparaison des performances des méthodes sur la base de visages FERET. Extrait de [PMRR00].	67
Tableau 4.1: Ensemble des données acquises lors de la campagne IV ² .	70
Tableau 4.2 : Séries de tests élaborées à partir des données collectées.	73
Tableau 4.3 Meilleurs scores de chaque algorithme pour le test 04a.	75
Tableau 4.4 Tableau récapitulatif des différents tests pour les évaluations du visage 2D et de la multimodalité réalisées.	77
Tableau 4.5 : Bases de données de visage formant le corpus de notre évaluation.	81
Tableau 5.1 : Tailles respectives des bases d'apprentissage, d'enrôlement et de test pour chacun des 4 experts.	87
Tableau 5.2 : Stratégie du choix des images pour les trois bases du premier expert.	87
Tableau 5.3 : Variation de la taille de la base d'enrôlement pour chacun des quatre experts.	89
Tableau 5.4 : Taux d'identification pour chaque vecteur caractéristique avec un ordre $n = 10$.	94
Tableau 5.5 : Dimensionnalité de certaines inerties de sous-espaces de visages et les taux d'identification relatifs.	96
Tableau 5.6 : résultats d'identification par les <i>Fisherfaces</i> avec les paramètres suggérés par Belhumeur et al. [BHK97] sur la base XM2VTS.	97
Tableau 5.7 : Résultats d'identification <i>Fisherfaces</i> pour différents sous-espaces de projection.	98
Tableau 5.8 : Relation entre les moments de Zernike et leurs ordres.	99
Tableau 5.9 : la taille g du vecteur caractéristique correspondant à une famille des moments de Zernike d'ordre n .	100
Tableau 5.10 : Comparaison des critères d'homogénéités et de séparabilité entre les classes calculé par les k -moyennes.	106
Tableau 5.11 : Taux de C-Identification avec un partitionnement en deux et en trois classes.	107
Tableau 5.12 : Tailles des vecteurs attributs et nombre de classes pour chaque type de descripteur.	115
Tableau 5.13 : Les différents modes de combinaison des méthodes de représentation des attributs faciaux entre le processus de partitionnement et celui d'identification finale.	115
Tableau 5.14 : Comparaison des temps d'exécution en seconde de l'approche classique et de notre approche d'identification pour l'ensemble des 400 images de test.	118
Tableau 6.1 : Nombre de personnes et d'instances constituant chaque sous base de notre corpus de travail.	122
Tableau 6.2 : dimensionnalité de quelques sous espaces de visage relativement à des inerties allant de 60% à 95% de l'espace de visage total.	124
Tableau 6.3 : Taille g de chaque descripteur des trois méthodes de représentation étudiées pour les deux phases d'identification et de clustering.	126

Tableau 6.4 : Résultats de réduction de la galerie par BIRCH avec un seuil d'absorption $T=0$	135
Tableau 6.5 : Résultats de réduction de la galerie par BIRCH avec un seuil d'absorption $T>0$	136
Tableau 6.6 : Nombre de classes K pour les différentes stratégies de classification.	136
Tableau 6.7 : Taux de pénétration R des différents classifieurs.	136
Tableau 6.8 : Tableau de contingence de deux partitions différentes de la galerie.	138
Tableau 6.9 : Test d'indépendance de χ^2 entre les partitions de différentes méthodes de représentation.	139
Tableau 6.10 : Test d'indépendance de χ^2 entre les partitions de différentes méthodes de classification.	139
Tableau 6.11 : Test d'indépendance de χ^2 entre les partitions de différentes mesures de similarités dans la classification des attributs <i>Eigenfaces</i>	139
Tableau 6.12 : Matrice de confusion entre le classifieur basé sur les attributs <i>Eigenfaces</i> (en lignes) et le classifieur basé sur les attributs <i>Fisherfaces</i> (en colonnes).	143
Tableau 6.13 : Nombre de personnes retenues entre les différentes paires de classifieurs.	143
Tableau 6.14 : Matrice de confusion entre le classifieur basé sur les attributs <i>Fisherfaces</i> et la classification BIRCH (en lignes) et le classifieur basé sur les moments de Zernike et le partitionnement par les K-moyennes (en colonnes).	144
Tableau 6.15 : Comparaison des temps de la recherche en secondes de l'approche classique et de notre approche d'identification indépendamment du temps requis pour l'extraction des attributs.	149
Tableau A.1 : Taille moyenne des données pour une personne	177

Introduction Générale

Contexte et motivations

Depuis les attentats du 11 septembre 2001 aux Etats-unis, et la médiatisation qui en a été faite, la sécurité est devenue une préoccupation internationale. Une des conséquences a été un meilleur contrôle du taux migratoire par les pays, notamment européens et américain, en s'assurant mieux de l'identité des voyageurs. Cette préoccupation de sécurité est d'autant plus aigüe dans notre temps moderne que les moyens de communication se multiplient. La nécessité de la protection civile d'une part et la lutte contre les fraudes et les crimes d'autre part, placent au centre un dispositif sécuritaire pour de nombreux domaines comme par exemple le transport, le secteur bancaire, les services publics, etc. Le dénominateur commun, est d'offrir des moyens simples, pratiques, fiables, pour vérifier ou identifier une personne, sans l'assistance d'une autre personne.

Afin de répondre à ces besoins liés à la sécurité, la biométrie se présente comme une technologie potentiellement puissante. En effet, les différents moyens biométriques visent à utiliser des caractéristiques comportementales et/ou physiologiques spécifiques à chaque personne. Empreintes digitales, numérisation de l'iris ou géométrie de la main, plusieurs de ces mesures ont été techniquement prouvées et commercialisées. Cependant, ils exigent, généralement, la coopération des utilisateurs et présentent ce caractère contraignant en obligeant les personnes à se plier aux exigences techniques du procédé. A la différence de ces modalités biométriques, la reconnaissance faciale ne nécessite pas théoriquement la coopération des participants. Elle apparaît donc comme une alternative très intéressante, à condition cependant qu'un taux de reconnaissance suffisant soit garanti. Ce caractère non invasif a été la raison majeure de l'intérêt grandissant de la reconnaissance faciale, de la part d'organisations de recherche publiques et privées.

Les premières études théoriques sur la reconnaissance faciale remontent au moins au début des années 1970. Le passage vers des produits commerciaux n'a reçu une impulsion décisive qu'à partir des années 1994-1996 en bonne partie grâce à la mise en oeuvre d'un programme d'évaluation international FERET (Face Recognition Technology), organisé par le ministère de la Défense américaine (*Department of Defense, DoD*). On dispose alors d'une base d'images de référence pour les tests. La fin du programme en 1996 a marqué l'apparition sur le marché des premiers produits commerciaux. Le vif intérêt du marché potentiellement gigantesque a fait éclore un grand nombre d'algorithmes de reconnaissance de visages ou de variantes, dont la plupart n'étaient même pas présents lors des tests d'évaluation FERET. Le DoD décida alors de lancer des programmes d'évaluation, FRVT (*Facial Recognition Vendor Test*) successivement en 2000, 2002 et en 2006 dont l'objectif était d'évaluer les performances des produits commerciaux.

Comme tout autre sujet de reconnaissance de formes, les problèmes liés à la reconnaissance de visages peuvent être caractérisés par les deux difficultés suivantes : (a) la variabilité intra-classes et (b) la similarité inter-classes. La première difficulté est d'abord liée aux variations

intrinsèques du visage produites par exemple par les changements d'expressions faciales, l'évolution de la physionomie du visage et les transformations du visage voulues ou subies. Cette variabilité englobe également des variations environnementales provoquées par exemple par les changements de conditions d'éclairage, l'orientation du visage de la personne ou l'occlusion d'un visage par des accessoires ou par d'autres objets. La seconde difficulté caractérise la similarité entre des classes d'individus différents. Ici, les apparences faciales de deux personnes différentes peuvent être similaires par exemple en raison des liens familiaux (jumeaux, jumelles, père/fils, mère/fille, etc.) ou plus généralement de l'appartenance à une même race.

Plusieurs de ces problématiques liées aux changements d'apparence d'individus dus à l'éclairage, à la posture, aux expressions faciales, au vieillissement, ... ont été étudiées au cours des campagnes d'évaluation FRVT. Ces évaluations ont permis de caractériser l'impact et de tracer les limites de chacun de ces facteurs. Ainsi, les résultats de ces campagnes, ont abouti à la conception de systèmes d'authentification performants. Par contre dans le cadre de l'identification plus générale, où l'on ne dispose d'aucune information a priori concernant l'identité, la plupart des systèmes connaissent une baisse de leurs performances dans des conditions réelles d'application. Cette problématique a été soulevée exclusivement par la campagne FRVT 2002. Nous reviendrons plus en détail sur cette problématique dans le premier chapitre.

Contribution et plan du manuscrit

Les travaux effectués dans le cadre de cette thèse portent sur l'identification automatique de visages dans des bases de données d'images numériques. L'objectif est de simplifier le déroulement de l'identification biométrique qui consiste à la recherche de l'identité requête parmi l'ensemble des individus de la base de données préenregistrée, appelée aussi galerie. En effet, le schéma d'identification classique est complexe et très coûteux notamment dans le cas de grandes bases de données biométriques. Le processus original que nous proposons ici a pour objectif de réduire la complexité et d'améliorer les performances en termes de temps de calcul et de taux d'identification.

Dans ce cadre biométrique, nous avons proposé une étape de classification non supervisée ou clustering des visages pour partitionner la base d'images en plusieurs sous ensembles cohérents et bien discriminés. Pour ce faire, l'algorithme de clustering vise à extraire, pour chaque visage, un ensemble de descripteurs, appelé signature, qui lui soit spécifique. Trois techniques de représentation faciales ont été développées dans le but d'extraire des informations différentes et complémentaires décrivant le visage humain : deux méthodes factorielles d'analyse multidimensionnelle et de projection des données (méthodes dites « *Eigenfaces* » et « *Fisherfaces* ») ainsi qu'une méthode d'extraction des moments géométriques de Zernike. Sur la base des différentes classes de signatures extraites pour chaque visage plusieurs méthodes de clustering sont mises en compétition afin de dégager la classification optimale qui conduit à une meilleure réduction de la galerie. Les méthodes retenues sont, soit de type « centres mobiles » telles que les K-moyennes et les centres mobiles de Forgy, ou de type « agglomérative » telle que la méthode de BIRCH. Sur la base de la dépendance des partitions générées, ces différentes stratégies classificatoires sont ensuite combinées suivant une architecture parallèle de manière à optimiser la réduction de l'espace de recherche à un sous ensemble de la base de données. Les clusters retenus in fine étant ceux pour lesquels la probabilité d'appartenance de l'identité requête est quasi certaine.

Le manuscrit de la thèse est organisé autour de six chapitres, de la manière suivante :

- Le chapitre 1 expose la problématique générale de la reconnaissance biométrique. Les mesures de performances couramment utilisées pour l'évaluation de tels systèmes sont présentées et les limitations rencontrées lors du passage à des bases de grande taille sont évoquées, en particulier pour l'identification. Ces limitations sont illustrées à partir des résultats des dernières campagnes d'évaluation en biométrie. Par exemple les trois meilleurs systèmes évalués dans le cadre de la campagne FRVT 2002 connaissent une baisse de 20 à 25% de leurs taux d'identification en passant d'une galerie de petite taille à une galerie de plusieurs milliers de visages. Pour finir un état de l'art de l'utilisation de stratégies de classification en biométrie est réalisé, permettant de mettre en avant l'objectif de ce travail, à savoir la réalisation d'une partition de la base de données avant le passage à l'identification finale.
- Pour reconnaître une personne à partir de son image, il faut passer par certaines étapes. Tout d'abord il faut détecter et/ou localiser le visage dans l'image. Le deuxième chapitre de cette thèse s'arrête sur cette étape de détection. Après un état de l'art de ces méthodes, deux approches concurrentes de localisation, l'une basée sur les moments de Zernike et l'autre sur les *Eigenfaces*, sont détaillées. Les résultats obtenus par chacune d'elles sur les données de la base XM2VTS sont exposés en utilisant un critère de qualité développé au laboratoire.
- Le chapitre 3 est consacré à un état de l'art sur la reconnaissance des visages. L'extraction des caractéristiques locales ou globales sont exposées ainsi que les difficultés inhérentes à de tels systèmes : changements d'expressions, de poses, d'éclairage, d'apparence au cours du temps,... Ce chapitre permet d'exposer les raisons nous ayant conduit à retenir trois méthodes pour le développement de notre approche future, à savoir « *Eigenfaces* », « *Fisherfaces* » et moments de Zernike.
- Le développement d'un système d'identification biométrique implique l'utilisation d'une base de données faciales pour la phase d'évaluation. Durant cette thèse, j'ai participé à toutes les phases de développement d'une base de données biométriques dans le cadre du projet IV² : Identification par l'Iris et le Visage via la Vidéo. Les 300 identités de la base IV², ajoutées aux données provenant d'autres bases Benchmark (XM2VTS, FERET, AR, ORL et CVL) nous ont permis de constituer un large corpus de travail de 1880 individus. La présentation de ces différentes bases d'images fait l'objet du chapitre 4.
- La présentation de notre approche originale de réduction de l'espace de recherche par classification non supervisée est exposée au chapitre 5. A partir des trois méthodes d'extraction d'attributs retenues au chapitre 3, nous mettons en œuvre une approche de partitionnement des données. Le principe de base de cette stratégie originale consiste à conduire l'identification en deux temps. Dans un premier temps on recherche, à partir de l'identité requête, un sous-ensemble de la base de données présentant une forte probabilité de ressemblance avec l'identité requête. Dans un deuxième temps, l'identification proprement dite ne sera conduite que dans cet ensemble réduit de visages. L'exposé de cette approche est réalisé sur la base XM2VTS en utilisant la méthode des K-moyennes pour l'étape de classification. L'analyse sémantique des classes obtenues permet d'illustrer l'intérêt de cette approche de classification non supervisée.
- Le dernier chapitre est consacré à la généralisation de notre stratégie. Généralisation en termes de données et en termes de méthodes de classification. En effet, outre les K-moyennes, nous explorons dans ce chapitre la méthode des centres mobiles de Forgy et celle de BIRCH. Après avoir analysé la réduction obtenue par chaque méthode, nous avons proposé une stratégie de combinaison de classifieurs permettant de valider les données retenues pour la recherche. Nous réalisons ainsi un gain sur ce grand corpus de travail en réduisant la taille de la galerie à 50% et en maintenant un taux de faux rejet

quasi nul. Des résultats en terme de taux d'identification et de temps de calcul sont comparés entre une approche d'identification directe et notre approche d'identification incluant l'étape de classification non-supervisée.

Chapitre I

Problématique de l'identification dans les bases de données biométriques

1.1 Introduction

Dans son environnement quotidien, un individu a besoin de s'identifier dans une multitude de contextes : pour entrer dans son immeuble ou accéder à son lieu de travail, pour retirer de l'argent à un distributeur ou payer en magasin, pour demander un service social... Autant de codes et de mots de passe à mémoriser et à protéger. Afin de développer les moyens de reconnaissance, la recherche connaît depuis quelques années un renouveau spectaculaire et manifeste un intérêt majeur aux données "biométriques", c'est-à-dire aux caractéristiques propres à chaque personne : sa voix, ses empreintes digitales, les traits de son visage, la forme de sa main, sa signature et même son ADN. Ainsi, les techniques basées sur la biométrie jouissent d'un engouement général favorisé par un phénomène de mode, principalement véhiculé par les films au cinéma et à la télévision. Cependant, plus récemment, l'augmentation de la fraude à l'identité a créé un besoin croissant de la technologie biométrique dans un certain nombre d'applications nécessitant un haut degré de sécurité : accès à des sites sensibles, surveillance d'aéroport [JBP99, RNJ06].

Il existe traditionnellement deux manières d'identifier une personne :

1. Méthodes basées sur une connaissance (knowledge-based). Cette connaissance correspond par exemple à un mot de passe pour ouvrir une session ou un code SIM pour un téléphone portable.
2. Méthodes basées sur une possession (token-based). Il peut s'agir d'une pièce d'identité, d'un badge, d'une clé ...

Ces deux modes d'identification peuvent être utilisés d'une manière complémentaire afin d'obtenir une sécurité accrue comme pour la carte bleue. Cependant elles ont leurs faiblesses respectives. Le mot de passe peut être oublié ou deviné par une autre personne. On estime ainsi qu'une personne sur quatre écrit son mot de passe sur sa carte bleue pour ne pas l'oublier [JHP00]. D'autre part, la pièce d'identité peut être perdue ou volée. La biométrie exploite à l'encontre de ces deux méthodes les caractéristiques d'une personne qu'elles soient innées comme les empreintes digitales ou acquises comme la signature. Ces caractéristiques sont attachées à chaque individu et ne souffrent donc pas des faiblesses des méthodes basées sur une connaissance ou une possession. En effet les caractéristiques biométriques ne peuvent être oubliées ou perdues. De plus, elles sont très difficiles à deviner, à voler, ou à dupliquer.

1.2 La reconnaissance biométrique

La reconnaissance biométrique fait référence à l'utilisation de différentes caractéristiques physiologiques (empreintes digitales, visage, rétine, iris, voix) et comportementales (manière de marcher, dynamique de la signature) [Fau05, Dau07, JRP04], appelées caractéristiques

biométriques, ou simplement biométrie, pour reconnaître automatiquement les individus. Toutes les caractéristiques biométriques d'une personne sont, en fait, une combinaison de caractéristiques physiologiques et comportementales. Par exemple, les empreintes digitales sont de nature physiologique, mais la façon dont un utilisateur présente son doigt au capteur dépend du comportement de la personne. Ainsi, l'entrée d'un système de reconnaissance biométrique est une combinaison de caractéristiques physiologiques et comportementales. De même, la voix est en partie déterminée par la structure biologique qui produit la parole chez un individu et aussi par la façon dont une personne parle. Souvent, une similitude peut être constatée entre les parents, les enfants, et les frères et sœurs dans leur voix, leur démarche, et même dans leur signature. Le même argument s'applique aux visages: les visages de jumeaux peuvent être très similaires à la naissance, mais au cours du développement, les visages changent en fonction du mode de vie et du comportement de la personne. 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 observés entre deux visages différents. Ceci se traduit par une baisse significative des performances de la reconnaissance automatique dans les conditions réelles pour le visage ainsi que pour la majorité des modalités biométriques. Ainsi la reconnaissance automatique des caractéristiques biométriques reste un problème complexe et non efficacement résolu.

1.3 Les caractéristiques biométriques

Les caractéristiques biométriques ne peuvent pas être facilement volées, falsifiées, ou partagées. Ainsi, elles sont plus fiables et sécurisées pour la reconnaissance de personne que les méthodes traditionnelles basées sur la connaissance ou la possession. Cependant ces caractéristiques physiques et comportementales doivent satisfaire plusieurs contraintes pour une grande fiabilité des systèmes biométriques. En effet, les objectifs de la reconnaissance biométrique sont la facilité d'utilisation par une reconnaissance sans carte ou code PIN, la sécurité accrue qui est traduite par la difficulté de contourner le contrôle d'accès ainsi que la plus grande performance quant à la précision et la rapidité des traitements. Ainsi, chaque caractéristique physiologique et/ou comportementale peut être utilisée comme une biométrie pour reconnaître une personne dès lors qu'elle répond à ces exigences [MMJP05, PD02]:

- *l'universalité*, ce qui signifie que chaque personne devrait avoir cette caractéristique biométrique ;
- *l'unicité*, qui indique que la caractéristique biométrique est représentative d'un et un seul individu;
- *la facilité de mesure*, qui se réfère à la facilité avec laquelle une personne peut interagir avec un système biométrique pendant qu'elle est identifiée ou authentifiée par ce système. Cette mesure est objective et dépend du type de dispositif utilisé pour capturer l'échantillon biométrique (caméra, scanner, etc.) ;
- *la performance*, qui se réfère principalement à la précision de la reconnaissance, à la vitesse d'exécution et à la robustesse face aux variations intrinsèques des caractéristiques biométriques, aux bruits et aux déformations des appareils de captures. La performance d'un système dépend aussi des ressources nécessaires pour atteindre la vitesse et la précision escomptées, ainsi que les facteurs opérationnels et environnementaux qui influent sur la vitesse et la précision de la reconnaissance;

Cependant, dans un système biométrique pratique, il existe un certain nombre d'autres paramètres qui doivent être examinés, notamment:

- L'acceptation par l'utilisateur qui reflète l'ampleur des inquiétudes et des objections que l'utilisation d'une technologie biométrique donnée tend à susciter. Dans certains pays, la reconnaissance faciale est mal acceptée ; dans d'autres pays, la reconnaissance des empreintes digitales a des connotations de criminalité. La mesure de l'acceptation reste très subjective et varie d'une personne à l'autre et d'un pays à l'autre, suivant le régime de protection des données en vigueur, le contexte culturel et les attentes personnelles des utilisateurs ;
- Le contournement, qui reflète combien il est facile de tromper le système par des méthodes frauduleuses.
- La permanence ou la stabilité qui est définie par la constance d'une caractéristique biométrique au cours du développement normal et du vieillissement d'une personne. En principe, plus une caractéristique est stable, moins il est nécessaire de mettre à jour les caractéristiques personnelles ou de réenregistrer la personne.

Un système biométrique pratique doit avoir une précision acceptable et une vitesse de reconnaissance raisonnable vis-à-vis des ressources requises, inoffensif pour les utilisateurs, accepté par la population, et suffisamment robuste à l'encontre des méthodes frauduleuses.

De nombreuses modalités biométriques sont utilisées dans diverses applications (cf. Figure 1.1). Chaque modalité biométrique a ses forces et ses faiblesses et le choix dépend généralement de l'application à traiter. Aucune modalité biométrique ne répond effectivement aux exigences de toutes les applications. Ainsi, l'utilisation de plusieurs modalités (multimodalités) biométriques pour l'identification d'un individu augmente aussi bien la précision que la couverture de la population tout en laissant décroître la vulnérabilité à l'imposture.

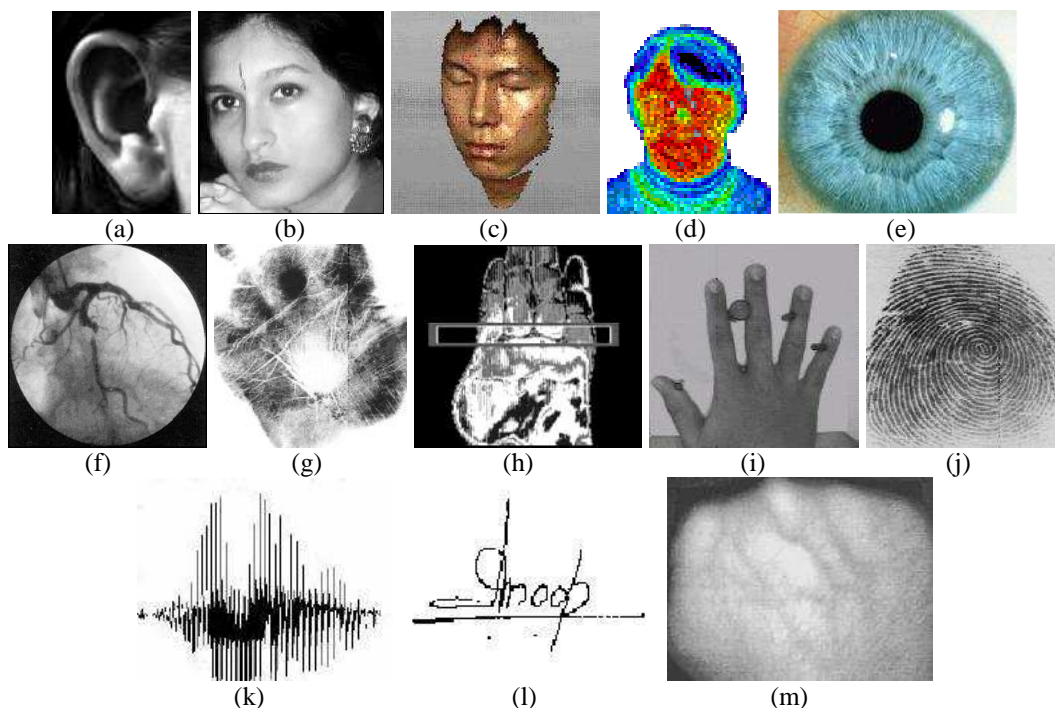


Figure 1.1 : Les principales caractéristiques biométriques pouvant servir comme moyen d'identification : a) forme de l'oreille, b) visage 2D, c) visage 3D, d) visage infrarouge, e) iris, f) rétine, g) empreinte de la main, h) thermogramme de la main, i) forme de la main, j) empreinte digitale, k) voix, l) signature et m) réseau veineux de la main [MMJP05].

La comparaison entre les différentes biométries permet de choisir une technologie en fonction des contraintes liées à l'application. En effet, chaque caractéristique (ou modalité) biométrique a ses forces et ses faiblesses, et faire correspondre un système biométrique spécifique à une application dépend du mode opérationnel de l'application et des caractéristiques biométriques choisies. En France le Club de la Sécurité des Systèmes d'Information Français [Clu04] a proposé une comparaison (avantages / inconvénients) des principales technologies biométriques en se basant sur la facilité ou l'ergonomie d'utilisation, la vulnérabilité aux attaques et aux contournements et la fiabilité relative à la précision et à l'efficacité de la reconnaissance (cf. tableau 1.1).

Techniques	Avantages	Inconvénients
Empreintes digitales	Coût, ergonomie moyenne, facilité de mise en place, taille du capteur	fiabilité des appareils de mesure, acceptabilité moyenne, possibilité d'attaques (rémanence de l'empreinte,...)
Forme de la main	Très ergonomique, bonne acceptabilité	Système encombrant, coût, perturbation possible par des blessures et l'authentification des membres d'une même famille, permanence des données
Visage 2D	Coût, peu encombrant, bonne acceptabilité	Jumeaux, psychologie, déguisement, vulnérabilité aux attaques
Rétine	Fiabilité, pérennité	Coût, acceptabilité faible, installation difficile
Iris	Fiabilité	Acceptabilité très faible, contrainte d'éclairage
Voix	Facilité	Vulnérable aux attaques
Signature	Ergonomie	Dépendant de l'état émotionnel de la personne, fiabilité
Frappe au clavier	Ergonomie	Dépendant de l'état physique de la personne

Tableau 1.1 : Avantages et inconvénients des technologies biométriques [Clu04].

1.4 Conception d'un système biométrique

Une question importante dans la conception d'un système biométrique est de déterminer la façon dont un individu est reconnu. En fonction du contexte de l'application, un système biométrique peut être soit un système de vérification, soit un système d'identification:

- un système de vérification vérifie l'identité d'une personne en comparant la signature biométrique extraite avec son propre modèle biométrique préenregistré dans le système. Il effectue une comparaison un à un (on note généralement 1:1) pour déterminer si l'identité revendiquée par l'individu est vraie. Un système d'authentification accepte ou rejette la demande d'un utilisateur : **suis-je ce que je prétends être ?**
- un système d'identification reconnaît un individu par une recherche sur l'ensemble de la base des modèles. Il effectue une comparaison un à plusieurs (on note généralement 1:N) pour établir l'identité de l'individu. Dans un système d'identification, l'utilisateur ne revendique pas une identité mais pose plutôt la question : **qui suis-je ?** Ainsi, le système doit établir l'identité d'un individu ou inférer que le sujet n'est pas enregistré dans la base de données du système.

Le terme authentification est fréquemment utilisé comme synonyme de vérification dans le domaine de la biométrie. D'autre part, la reconnaissance biométrique englobe le plus souvent les deux modes d'authentification et d'identification. Cependant dans un cadre propre à l'identification biométrique, le terme reconnaissance fait parfois référence seulement à la présence ou non d'une personne dans la base de données du système sans chercher à l'identifier. Dans ce cadre de recherche, peu de travaux permettent de déduire l'appartenance d'une personne dans une base de données indépendamment de son identification [O'DV94, O'T04]. L'identification est la tâche la plus communément effectuée. Le processus de recherche utilisé par la plupart des algorithmes retourne une identité spécifique plutôt qu'un ensemble ou une partition d'appartenance [LJ04].

Les schémas d'un système de vérification et d'un système d'identification sont illustrés dans la figure 1.2; le processus d'enrôlement, qui est commun à ces deux tâches est également illustré. Le module d'enrôlement correspond à l'enregistrement biométrique des individus dans la base de données du système. Pendant la phase d'enrôlement, la caractéristique biométrique d'un individu est capturée par un lecteur biométrique pour produire une représentation numérique

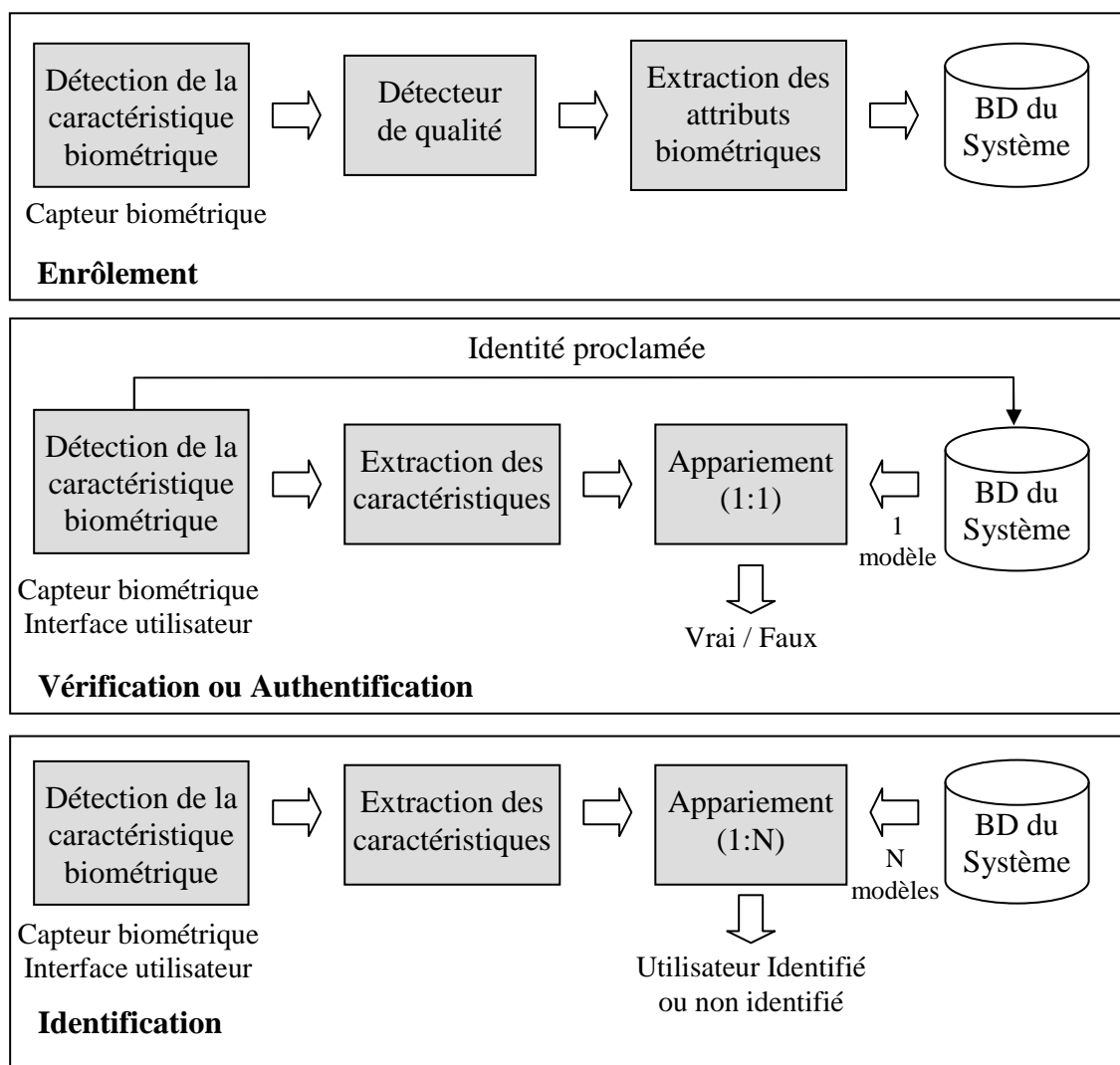


Figure 1.2 : Schéma de fonctionnement d'un système biométrique. Diagrammes des processus d'enrôlement, de vérification et d'identification.

de cette biométrie. Un contrôle de qualité est généralement effectué pour s'assurer que la prise de l'échantillon est effectuée de manière fiable et pour garantir une bonne qualité de l'acquisition [JKL03]. Afin de faciliter l'appariement, la représentation numérique extraite par le capteur est généralement traitée par une fonction d'extraction pour générer une signature compacte et expressive, appelé aussi modèle. Selon l'application, le modèle peut être stocké dans la base de données centrale du système biométrique ou être enregistré sur une carte magnétique ou carte à puce délivrée à la personne.

Le processus de vérification est établi pour authentifier une personne se présentant à un point d'accès. Durant la phase de vérification, le nom de l'utilisateur ou le code PIN (*Personal Identification Number*) est entré par l'intermédiaire d'un clavier. Le lecteur biométrique capte la biométrie de l'individu et la convertit à un format numérique, qui est traité ultérieurement par la fonction d'extraction pour produire une signature compacte qui représente l'identité de l'utilisateur. Cette signature numérique (appelée aussi vecteur caractéristique) est présentée à la fonction d'appariement, qui la compare avec le modèle proclamé par l'utilisateur et qui est extrait de la base de données (BD) du système à partir de son code PIN.

Le processus d'identification ne nécessite pas de code PIN. Le système compare la signature biométrique saisie avec les modèles de tous les utilisateurs dans la base de données du système ; la sortie est généralement l'identité d'un utilisateur enrôlé ou un message du genre «utilisateur non Identifié» si l'individu n'est pas enregistré dans la base de données du système. L'identification dans les grandes bases de données est ainsi coûteuse en termes de complexité et efficacité de calcul. Les techniques de classement et d'indexation s'imposent pour limiter le nombre de modèles qui doivent être comparés avec la signature de l'utilisateur.

La procédure d'enrôlement est opérée généralement d'une façon semi automatique suivant un mode hors ligne. En effet, l'enrôlement peut être supervisé par un expert ou un agent de police qui peut orienter le processus d'acquisition quand un criminel est enrôlé, par exemple. D'autre part, les processus d'authentification et d'identification s'opèrent en ligne. Ces procédures de reconnaissance en ligne doivent s'exécuter rapidement car une réponse immédiate est imposée dans la majorité de ces applications.

1.4.1 Mode de reconnaissance

Suivant l'application traitée, un système biométrique peut fonctionner suivant un mode de reconnaissance positif ou négatif:

- Dans une reconnaissance *positive*, le système permet de déterminer si la personne est celle qu'elle prétend être, de manière implicite ou explicite. Le but d'une reconnaissance positive est d'empêcher plusieurs personnes d'utiliser la même identité. Par exemple, si une personne unique est autorisée à entrer dans une certaine zone sécurisée, le système lui donnera exclusivement l'accès. Si le système arrive à faire correspondre la signature enregistrée de cette personne à la signature extraite de la biométrie acquise, cela correspond à une acceptation, et sinon à un rejet;
- Pour une demande de reconnaissance *négative*, le système permet de déterminer si la personne correspond à ce qu'elle nie être de manière explicite ou implicite. L'objet de la reconnaissance est négatif pour empêcher une seule personne d'utiliser de multiples identités. Par exemple, pour une application de prestations sociales, le système enregistre dans sa base de données les personnes qui ont déjà bénéficié de prestations. Si une personne malhonnête souhaite recevoir les prestations une deuxième fois en se proclamant

comme étant une tierce personne, le système doit vérifier si celle-ci correspond à un des bénéficiaires enregistrés dans la base de données. Si le système arrive à faire correspondre la signature extraite de la biométrie acquise de cette personne avec une des signatures de la base de données, cela correspond à un rejet, et sinon à une acceptation.

Notons que les méthodes traditionnelles d'authentification, telles que les mots de passe, les codes PIN, les clés, les jetons peuvent être utilisées uniquement pour une reconnaissance positive. La reconnaissance négative, quant à elle, ne peut être établie que par la biométrie. En outre, la reconnaissance positive peut fonctionner aussi bien en mode de vérification ou d'identification, mais les applications qui nécessitent une reconnaissance négative ne peuvent être déployées qu'en mode d'identification: en fait, le système doit parcourir la totalité de ses bases de données pour prouver que les données saisies ne sont pas présentes.

1.4.2 Applications et enjeux

Dans cette section, nous présentons les principales applications de la technologie de la reconnaissance biométrique, ainsi que leurs enjeux. Le tableau 1.2 répertorie ces applications ainsi que leurs principales caractéristiques. Le sigle « BD » désigne la base de données qui contient les signatures des personnes enregistrées ou enrôlées dans le système. La communauté de la reconnaissance faciale et biométrique en général a défini une fourchette évaluative en termes de nombre de signatures et d'individus [PGM04]. En considérant en moyenne trois signatures par personne enregistrées, la taille de la BD est qualifiée de petite si le nombre de signatures n'excède pas 100 ce qui correspond au plus à 33 individus. On parle d'une taille moyenne pour un nombre d'individus inférieur à 330 environ et de grande taille pour un nombre de personne avoisinant les 3300. Ensuite, viennent les systèmes à très grande échelle qui considèrent une base de l'ordre de 33000 individus ce qui correspond à une centaine de milliers de signatures.

Les applications biométriques peuvent être déployées en monde fermé (F) ou en monde ouvert (O). En monde fermé, tout visage requête est supposé être enregistré dans la base de données du système, alors qu'en monde ouvert, des visages de personnes inconnues peuvent être présentés au système.

La reconnaissance peut consister en une tâche d'identification (Id), ou d'authentification (Au). En considérant les images faciales comme caractéristique biométrique par exemple, identifier un visage, c'est lui assigner une identité sans prendre en compte aucune information à priori sur la classe d'appartenance présumée. En revanche, dans le contexte de l'authentification, toute personne se présentant au système revendique une certaine identité. Le processus consiste à vérifier qu'il s'agit bien de cette personne. D'autre part, comme nous l'avons vu dans la section précédente, un système peut fonctionner suivant un mode de reconnaissance positif (+) ou négatif (-).

Le niveau de sécurité de l'application est directement lié aux conséquences d'une mauvaise reconnaissance. Par exemple, on pourra considérer que si un système conçu dans un but de divertissement effectue une mauvaise reconnaissance, la portée de cette erreur sera moindre que si celle-ci avait amené à une usurpation d'identité pour une transaction financière.

Le *contrôle parental* consiste aux applications de la biométrie dans un cadre familial, pour l'accès à un poste informatique ou à la télévision, pendant un certain moment de la journée,

par exemple. S'agissant de solutions biométriques dans un cadre familial, on peut considérer que le monde est fermé et la base de données de petite taille.

Catégorie d'application	Tâche		mode	Taille de BD	Niveau de sécurité	Monde
	Id	Au				
Contrôle parental	X		+	petite	moyen	F
Contrôle d'accès		X	+	petite / moyenne	haut	O
Sécurisation des transactions		X	+	Très grande	haut	O
Indexation	X		+	grande	moyen	O
Surveillance / aide à la décision	X		-	moyenne/grande	haut	O
Interaction Homme Machine (IHM)	X		+	petite	faible	F

Tableau 1.2 : Les différentes catégories d'applications de reconnaissances biométriques vis à vis de la tâche et du mode de reconnaissance, la taille de la base de données, le niveau de sécurité, ainsi que le contexte [Vis05].

Le *contrôle d'accès* à son poste de travail, à un site sécurisé ou à sa session personnelle sur un système informatique sont typiquement des applications d'authentification en monde ouvert puisque des imposteurs peuvent tenter d'accéder à la zone protégée. Le nombre de personnes autorisées d'accès étant limité, la taille de la BD est petite à moyenne. Il est possible d'amener l'utilisateur à coopérer, par exemple en lui demandant de placer sa tête ou sa main à un endroit précis ce qui peut simplifier la tâche de la reconnaissance.

La *sécurisation des transactions* possède les mêmes caractéristiques que les applications de contrôle d'accès, sauf que le nombre de personnes autorisées est généralement très grand.

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* et de recherche parmi ces documents. La recherche d'identité dans un document multimédia est conçue généralement dans le but de déterminer si une personne connue est présente dans un fichier particulier ce qui peut aider à accéder à d'autres types d'informations. Les applications d'indexation sont nécessairement effectuées *a posteriori* de la prise de vue et donc elles reposent généralement sur une tâche d'identification hors ligne.

Les applications de *surveillance* déterminent secrètement et discrètement si une personne dans un lieu public appartient à une liste recherchée (*watch-list*) d'identités. On se place alors dans le cadre d'une tâche d'identification en mode négatif. Il est aussi nécessaire 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. 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 rendant la constitution de la base de données assez aisée. Ces applications de sécurité renferment

généralement des bases de données dont la taille est moyenne à grande. On distingue, dans cette catégorie aussi, les applications d'*aide à la décision* où on analyse *a posteriori* le contenu des applications de surveillance dont le temps de réponse doit être assez rapide comme par exemple la recherche de comportement à risque dans une rame de métro. D'autre part, la phase d'enrôlement n'est généralement pas bien effectuée et les images détectées peuvent être de mauvaise qualité et de faible résolution constituant des bases de données illégales. Nous notons finalement, que la surveillance et l'identification à grande échelle ne peuvent pas être accomplies par les techniques traditionnelles basées sur la connaissance ou la possession mais seulement par des techniques biométriques.

La reconnaissance de personne est considérée de nos jours comme une étape préliminaire et critique pour la construction de machines intelligentes qui fonctionnent dans l'environnement humain. L'objectif des systèmes d'*Interaction Homme Machine (IHM)* est d'analyser les images et les signaux biométriques, comme les images de visages, pour interagir avec des moyens non conventionnels. Par exemple, déplacer un curseur avec les yeux ou interagir à travers les émotions du visage...

1.4.3 Applications visées

Dans cette thèse, nous nous intéressons plus spécifiquement à la tâche d'identification. Notre système développé en monde fermé peut aussi être utilisé dans le cadre d'une application d'identification en monde ouvert. Nous cherchons à concevoir un système fiable pour des tailles de base de données moyennes à très grandes c'est-à-dire allant de quelques centaines à plusieurs milliers d'individus. Nous ne visons aucune des applications citées ci-dessus en particulier. Néanmoins, le développement d'un tel outil nous ouvre les domaines de la vidéo surveillance et de l'aide à la décision dans le contexte de la sécurité.

1.5 Mesure des performances des systèmes de reconnaissance biométriques

L'évaluation d'un système de reconnaissance est réalisée principalement en mesurant deux critères : la précision et l'efficacité (vitesse d'exécution). D'autres mesures de performances telles que l'échec à l'acquisition ou à l'enrôlement, le volume de données stocké pour chaque individu ainsi que le coût des matériels et des logiciels, sont des facteurs à prendre en compte pour la mise en place d'un système biométrique [PMWP00, OECD05]. Nous nous concentrons dans ce qui suit sur les deux premiers critères à savoir la précision et l'efficacité dans le cadre des systèmes d'authentification pour les étendre aux systèmes d'identification.

La réponse d'un système de reconnaissance biométrique est généralement un score de similarité s (dans l'intervalle $[0,1]$), entre la signature de la biométrie acquise et le modèle enregistré dans la base de données. Plus le score est proche de un, plus le système est sûr que les deux images proviennent de la même personne. Plus le score est proche de 0, moins le système est confiant dans le fait que les deux images proviennent de la même personne. La décision du système est régie par un seuil t : les échantillons biométriques qui génèrent des scores supérieurs à t sont appariés et la conclusion d'appartenance à la même personne est prise. Par contre, les échantillons qui génèrent des scores inférieurs à t sont non appariés entraînant la conclusion qu'ils proviennent de deux personnes différentes.

Deux types d'erreurs peuvent être faites dans un système biométrique :

- Faux appariement (*false match*) ; le système biométrique apparie des échantillons ne provenant pas du même utilisateur.
- Faux non-appariement (*false non-match*) ; le système biométrique ne réussit pas à appairer des échantillons provenant du même utilisateur.

Il est à noter que ces deux types d'erreurs sont souvent désignés comme fausse acceptation et faux rejet respectivement. Nous proposons de distinguer entre ces terminologies suivant le mode de reconnaissance positif et le mode négatif (voir la section 1.4). Dans les systèmes de reconnaissance positifs (par exemple, un système de contrôle d'accès) le faux appariement correspond à la fausse acceptation d'un imposteur, alors qu'un faux non-appariement correspond à un faux rejet d'un utilisateur légitime. D'autre part, dans un mode de reconnaissance négatif (par exemple, le système, cité plus haut, empêchant les utilisateurs d'obtenir des prestations sociales sous de fausses identités), un faux appariement se traduit par le rejet d'une demande légitime, alors que l'acceptation erronée d'un imposteur implique un faux non-appariement. Ainsi, la notation que nous proposons "faux appariement / faux non-appariement", correspondant plus fidèlement aux termes anglophone « *false match / false non-match* », est indépendante d'une application particulière et est préférable à celle de "fausse acceptation / faux rejet". Toutefois, l'utilisation du taux de fausse acceptation (*FAR*) et les taux de faux rejet (*FRR*) est largement utilisée dans la communauté de la recherche ainsi que dans l'environnement commercial. Pour des raisons de simplicité et de lisibilité de ce manuscrit, nous avons choisi d'utiliser ces derniers termes (*FAR* et *FRR*) comme mesures de performances des systèmes de reconnaissance biométriques. Nous détaillons dans ce qui suit les erreurs et les moyens d'évaluation des systèmes d'authentification pour les étendre ensuite aux systèmes d'identification.

1.5.1 Erreur des systèmes d'authentification

Nous formulons le problème de l'authentification biométrique comme suit. Nous désignons par T le modèle biométrique stocké d'une personne et par I les données acquises par le système biométrique. Ainsi, les deux hypothèses qui peuvent être posées sont:

H_0 : $I \neq T$, la signature biométrique calculée et le modèle enregistré dans la base de données ne proviennent pas de la même personne;

H_1 : $I = T$, la signature biométrique calculée et le modèle enregistré dans la base de données proviennent de la même personne.

Les décisions associées à H_0 et H_1 sont :

D_0 : la personne n'est pas celle qu'elle prétend être;

D_1 : la personne est celle qu'elle prétend être.

L'authentification consiste à appairer T et I en utilisant une mesure de similarité $s(T, I)$. D_0 est décidée dans le cas où le score d'appariement est inférieur à un seuil t , et D_1 est décidée autrement. Le taux de fausses acceptations (*FAR*) est la probabilité que D_1 soit décidée alors que H_0 est vraie. Et le taux de faux rejets (*FRR*) est la probabilité que D_0 soit décidée alors que H_1 est vraie.

Pour évaluer la précision d'un système biométrique, on doit calculer des scores à partir d'échantillons biométriques appartenant à une même personne, et des scores issus d'échantillons biométriques de différentes personnes. La figure 1.3 illustre l'estimation des taux *FAR* et *FRR* suivant les distributions des imposteurs et des personnes légitimes données par les équations 1.1 et 1.2 [MMJP05].

$$FAR = \int_0^t p(s | H_1 = \text{vraie}) ds \quad (1.1)$$

$$FRR = \int_t^1 p(s | H_0 = \text{vraie}) ds \quad (1.2)$$

La distribution $p(s|H_1=\text{vraie})$ des scores issus d'échantillons biométriques appartenant à une même personne est appelée distribution des personnes légitimes. La distribution $p(s|H_0=\text{vraie})$ de scores issus d'échantillons biométriques de différentes personnes est appelée distribution des imposteurs. On voit que si l'on fait varier le seuil t , on va modifier les valeurs respectives de FAR et FRR .

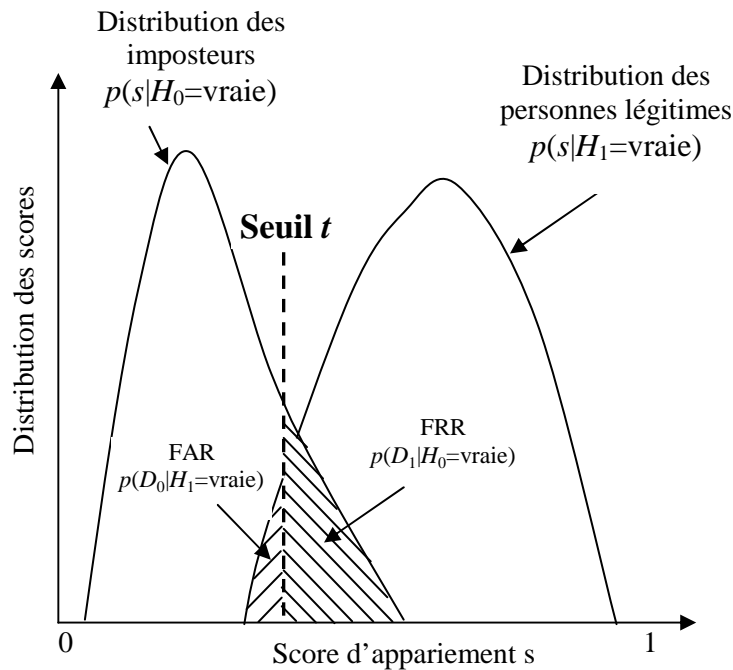
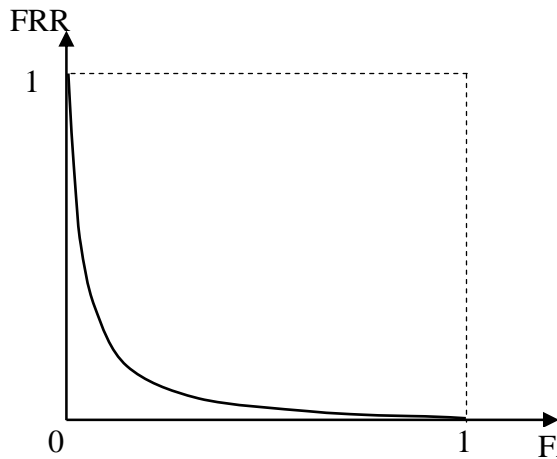
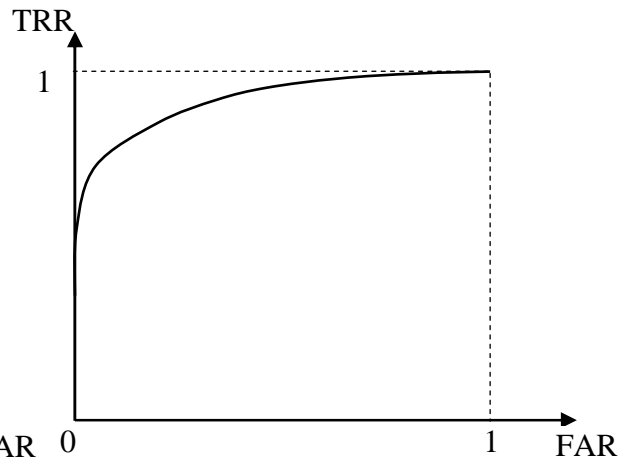


Figure 1.3 : La distribution des scores des personnes légitimes et des imposteurs et les taux d'erreurs, FAR et FRR , pour un seuil donné t .

Un compromis est à rechercher entre FAR et FRR dans chaque système biométrique [GMM97]. En effet, FAR et FRR sont fonctions du seuil de décision t du système et, plus rigoureusement, elles devraient être notées comme $FAR(t)$ et $FRR(t)$, respectivement. Leurs courbes respectives (voir figure 1.4) permettent d'estimer les performances du système pour différentes valeurs du seuil de décision. La représentation du FRR en fonction du FAR pour les différents points opérationnels du seuil t , permet de tracer la courbe DET « Detection Error Tradeoff » largement répandue pour l'évaluation de tels systèmes d'authentification. La courbe ROC « Receiver Operating Characteristics » est une autre représentation, pas moins connue, qui trace le taux de rejet justifié (*True Reject Rate* : TRR) en fonction du FAR . Les allures de ces deux courbes sont illustrées par les figures 1.4 et 1.5.



...Figure 1.4 : Courbe DET.



.....Figure 1.5 : Courbe ROC.

Le réglage du seuil de décision doit être fait en fonction du niveau de sécurité de l'application visée et des performances intrinsèques de la modalité biométrique utilisée. Nous présentons ci-dessous quelques points de fonctionnement qui résument la précision des systèmes de vérification.

- *Le taux d'équivalence des erreurs* (Equal Error Rate : *EER*) correspond au taux d'erreur pour lequel le taux de fausse acceptation et le taux de faux rejet sont identiques: $FAR(t) = FRR(t)$ (voir la figure 1.6). Bien que le *EER* soit un indicateur important, dans la pratique, un système biométrique est rarement évalué à ce point de fonctionnement, mais l'est souvent par un seuil de décision plus rigoureux placé pour réduire le *FAR* (par exemple à 1% ou 0,1%) malgré une élévation de *FRR* ou inversement.
- *ZeroFRR* correspondant à la valeur minimale de $FAR(t)$ pour laquelle aucun faux rejet ne se produit.
- *ZeroFAR* correspondant à la valeur minimale de $FRR(t)$ pour laquelle aucune fausse acceptation ne se produit.

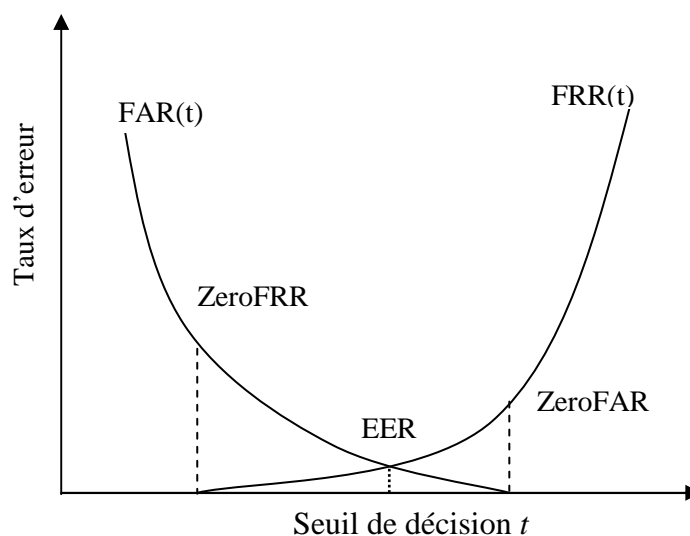


Figure 1.6 : Exemple des courbes $FAR(t)$ et $FRR(t)$, où les points correspondant aux *EER*, *ZeroFRR* et *ZeroFAR* sont illustrés.

1.5.2 Evaluation des systèmes d'identification

Le calcul des performances d'un système d'identification biométrique requiert trois ensembles d'images. Le premier est la galerie \mathcal{G} qui contient les échantillons biométriques des personnes connues par le système. Les deux autres sont des ensembles de test (*probe sets*) qui contiennent des échantillons biométriques qui sont présentés au système pour leur identification. Le premier ensemble de test $\mathcal{P}_{\mathcal{G}}$ (ayant l'initiale des termes anglais *probe* et *gallery*) contient des échantillons biométriques de personnes appartenant à la galerie \mathcal{G} . Ces échantillons de test sont différents de ceux de la galerie. L'identification en monde fermé requiert cet ensemble de test pour identifier une personne dans la galerie ($I_d \in \mathcal{P}_{\mathcal{G}}$ et $I_d \in \mathcal{G}$). L'autre ensemble de test $\mathcal{P}_{\mathcal{N}}$ contient des échantillons biométriques de personnes n'appartenant pas à la galerie ($I_d \in \mathcal{P}_{\mathcal{N}}$ et $I_d \notin \mathcal{G}$). Il s'agit ici du cas plus générique de l'identification en monde ouvert.

1.5.2.1 En monde fermé

En monde fermé, le taux d'identification est la mesure la plus couramment utilisée mais il n'est pas toujours suffisant. En effet, en cas d'erreur, il peut être utile de savoir si le bon choix se trouve dans les N premières réponses. On trace alors le score cumulé (*cumulative match score* - *CMS*) qui représente la probabilité que le bon choix se trouve parmi les premières réponses (cf. figure 1.7).

Dans le cas où il existe plusieurs signatures pour chaque individu dans la base de données, les mesures classiques des systèmes de recherche dans une base de données (*database retrieval system*) peuvent être utilisées. La précision (*precision*) est le rapport entre le nombre de modèles correctement retrouvés par le système dans la base de données et le nombre total de modèles retrouvés. Le rappel (*recall*) est le rapport entre le nombre de modèles correctement retrouvés dans la base de données et le nombre total de modèles qui auraient dû être retrouvés.

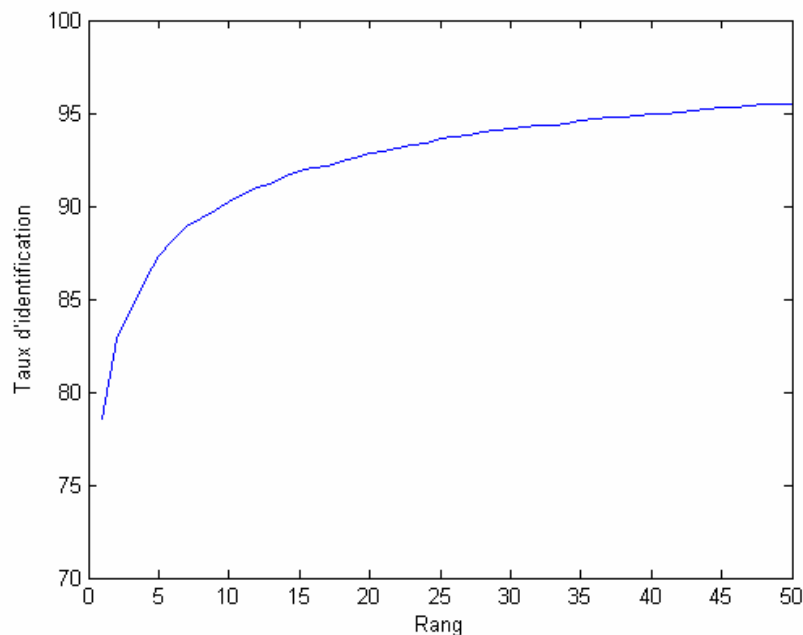


Figure 1.7 : Exemple de courbe de scores cumulés (CMS).

Comment peut-on étendre les définitions des erreurs présentées pour la vérification (FAR et FRR) à l'identification biométrique? Avec quelques hypothèses de simplification, l'estimation des performances en mode d'identification peuvent être inférées par les estimations d'erreur en mode de vérification.

Supposons qu'un seul modèle¹ (ou bien signature) pour chaque utilisateur est présent dans la galerie G et qu'aucun mécanisme d'indexation n'est disponible (c.-à-d., que toute la galerie contenant les N modèles doit être parcourue pour la recherche). Si le FRR et le FAR indiquent les taux de faux rejet et de fausse acceptation pendant la vérification, alors les taux correspondants en mode d'identification, FRR_N et FAR_N , sont donnés par [MPCG05, MCG05] :

- $FRR_N = FRR$; en effet, la probabilité de ne pas faire correspondre la signature extraite d'un utilisateur contre son modèle enregistré dans la galerie est la même qu'en mode de vérification;
- $FAR_N = 1 - (1 - FAR)^N$; en effet, une fausse acceptation se produit quand la signature de l'utilisateur correspond à tort à un ou à plusieurs modèles de la galerie. Le FAR_N est alors calculé comme "un" dont on soustrait la probabilité qu'aucune fausse acceptation n'est produite avec tous les modèles de la galerie. Dans l'expression ci-dessus $(1 - FAR)$ est la probabilité que la signature de l'utilisateur ne soit pas appariée (à tort) avec un seul modèle de la galerie, et $(1 - FAR)^N$ est la probabilité qu'elle ne soit pas appariée (à tort) à aucun des N modèles enregistrés dans la galerie. Plus le FAR est petit, plus l'expression ci-dessus peut être approchée par :

$$FAR_N \approx N \cdot FAR \quad (1.3)$$

Nous pouvons constater que la probabilité de fausse acceptation augmente linéairement avec la taille de la galerie.

La mesure des taux de réponse des systèmes d'identification est peu analysée (développée) dans la bibliographie [CMM00]. Pour combler cette lacune et par analogie avec les systèmes d'authentification nous développons dans ce qui suit la relation entre le taux d'identification (*Identification Rate* : IR) pour les systèmes d'identification d'une part et les taux de fausse acceptation (FAR) et de faux rejet (FRR) d'autre part.

Soit les taux suivants où le système réussit sa tâche de prise de décision :

- TAR le taux d'acceptation justifiée (*True Acceptance Rate* : TAR)
- TRR le taux de rejet justifié (*True Reject Rate* : TRR).

L'acceptation justifiée et le taux de faux rejet sont calculés pour une personne pré-enregistrée dans la base de données du système et autorisée à son accès. D'autre part, un imposteur ou une personne non autorisée engendre un rejet justifié ou une acceptation erronée comme décision du système. Ainsi, on peut écrire les relations suivantes :

¹ Les termes « modèle » ou « signature » biométriques sont exprimés comme étant un vecteur de caractéristiques ou d'attributs (*feature vector*) dans le problème de la reconnaissance biométrique et dans le problème plus générique de la reconnaissance de formes. Cependant, un modèle biométrique constitue généralement une représentation plus explicite qu'une signature biométrique.

$$TAR = 1 - FRR \quad (1.4)$$

$$TRR = 1 - FAR \quad (1.5)$$

Supposons que chaque identité enregistrée dans la galerie est représentée par une seule signature extraite à partir d'une ou de plusieurs modalités biométriques. Un système d'identification compare la signature extraite de la personne à identifier avec l'ensemble des N signatures de la galerie et retourne la plus proche signature correspondante à l'identité retrouvée. Pour chaque requête d'identification, il existe une plage de seuil $\delta t \subset [0,1]$ qui nous permet d'avoir $N-1$ rejet avec un score d'appariement inférieur à t , et une seule acceptation avec un score supérieur à t . Nous supposons que ces plages de seuils δt se chevauchent ou bien qu'elles sont assez proches pour l'ensemble des images de test \mathcal{P}_G . Il existe, alors, un certain seuil de décision τ qui nous permet d'avoir $N-1$ rejets et une seule acceptation pour l'ensemble des images de test. Nous pouvons alors écrire :

$$\begin{aligned} IR &= TAR \cdot TRR^{N-1} \\ &= (1 - FRR) \cdot (1 - FAR)^{N-1} \end{aligned} \quad (1.6)$$

En supposons que le FRR est très petit, nous pouvons écrire :

$$\begin{aligned} IR &\approx (1 - FRR) \cdot (1 - (N - 1) \cdot FAR) \\ &\approx 1 - (N - 1) \cdot FAR - FRR + (N - 1) \cdot FAR \cdot FRR \end{aligned} \quad (1.7)$$

d'où le taux d'erreur d'identification suivant :

$$\begin{aligned} \text{err} &\approx 1 - IR \\ &\approx (N - 1) \cdot FAR + FRR - (N - 1) \cdot FAR \cdot FRR \end{aligned} \quad (1.8)$$

En négligeant le terme $(N - 1) \cdot FAR \cdot FRR$ devant le terme $(N - 1) \cdot FAR$, nous obtenons :

$$\text{err} \approx (N - 1) \cdot FAR + FRR \quad (1.9)$$

1.5.2.2 En monde ouvert

Les performances pour un système d'identification en monde ouvert sont caractérisées par deux mesures de performances : le taux d'identification et le taux de fausse alarme. Dans le cas où l'identité recherchée se trouve dans la galerie, le taux d'identification est la fraction des images de test dans \mathcal{P}_G qui est correctement détectée et identifiée. Cette performance englobe, ainsi, les erreurs relatives à la détection et à l'identification dans le système fermé. Dans l'autre cas où l'échantillon de test n'appartient pas à une identité de la galerie (i.e. $\in \mathcal{P}_N$), le taux de fausse alarme (*False Alarm Rate*) correspond à la fraction des images de test dans \mathcal{P}_N qui est faussement détectée. Ces personnes sont référées comme étant des imposteurs dans les systèmes d'authentification.

1.5.3 Les intervalles de confiance

Les bases de données utilisées pour évaluer les performances des systèmes biométriques sont toujours de tailles réduites et jamais totalement représentatives de la population globale. En plus, les protocoles d'évaluation qui nécessitent le découpage de la base en deux sous-ensembles (apprentissage et test) ne permettent pas d'avoir une valeur significative de performance avec une seule valeur de taux d'erreur.

Une solution pour avoir un intervalle de confiance pour l'estimation des taux d'erreur est de répéter un certain nombre de fois le découpage en base de test et base d'apprentissage afin de pouvoir estimer l'erreur en généralisation. Une autre solution s'appuie sur des méthodes statistiques pour estimer un intervalle de confiance sur les taux d'erreurs [BRP04]. La mesure la plus utilisée estime les intervalles de confiance par une méthode paramétrique. Nous avons utilisé cette méthode pour la mesure des taux d'erreurs et essentiellement les taux d'identification (IR) dans cette thèse. Cette méthode impose l'hypothèse que les scores sont issus d'une loi normale dont on peut calculer la moyenne et la variance. Ainsi l'intervalle de confiance peut être déterminé grâce aux percentiles de la distribution normale. Par exemple l'intervalle de confiance à 90% pour IR correspond au percentile à 90% centré de la loi normale :

$$IR \in [\overline{IR} - 1,645\sigma, \overline{IR} + 1,645\sigma] \quad (1.10)$$

\overline{IR} est la moyenne du taux d'identification et σ est la variance de la loi normale associée à la distribution des scores d'identification.

Pour des intervalles de confiance à 95%, il suffit de remplacer 1,645 par 1,96.

1.6 Position du problème d'identification

Malgré la prolifération des bases de données à grande échelle, les effets sur les performances en utilisant ces bases de données ont été négligés. En effet, plusieurs applications, en particulier celles utilisées pour la sécurité, emploient des bases de données de grande échelle. Le facteur d'échelle, ou « scalabilité », a le potentiel d'affecter d'autant plus les performances des algorithmes que la taille des bases de données est grande.

Nous pouvons remarquer, d'après l'estimation des erreurs précédemment établie, que les systèmes d'identification biométrique (appariement 1:N) souffrent beaucoup de l'augmentation de la taille de la base de données. Généralement, la complexité de calcul est perçue comme le plus grand problème en développant les systèmes d'identification alors que la précision se dégrade plus vite que la vitesse d'exécution. Nous pouvons certainement trouver des algorithmes et des architectures de calcul plus rapides capables d'effectuer l'identification en quelques secondes pour une base de données conséquente. Par contre, un système qui a des performances acceptables pendant la vérification (appariement 1:1) peut devenir inutile une fois utilisé pour l'identification dans une grande base de données, et ceci à cause du facteur d'échelle. Par exemple, supposons qu'un système qui emploie une technique d'authentification performante donne un taux de fausses acceptations FAR de l'ordre de 10^{-5} (c'est-à-dire juste un faux appariement pour 100.000 vérifications). Si on étend cette technique de reconnaissance à un système d'identification avec une base de données de l'ordre de 10^4 utilisateurs (i.e. $N = 10.000$), la probabilité d'un faux appariement d'un modèle de la base

avec celui d'un imposteur est $FAR_N = 10\%$. Ceci suggère qu'un imposteur a une grande chance de tromper la sécurité du système en tentant un accès à plusieurs reprises ou en utilisant par exemple les 10 empreintes de ses deux mains pour un système qui utilise cette modalité.

1.6.1 Travaux d'évaluation large échelle

Plusieurs tests d'évaluation ont été élaborés, en majorité par le gouvernement Américain, depuis le début de cette décennie pour évaluer des systèmes biométriques proposés par des compagnies commerciales et des institutions académiques de recherche. Citons entre autre :

- La *Fingerprint Verification Competition* (FVC) organisée en 2000, 2002, 2004 et 2006 mesure les performances des systèmes d'authentification par empreintes digitales [MAI02].
- La *Speaker Recognition Evaluation* (SRE) coordonnée par le *National Institute of Standards and Technologies* (NIST) est une campagne d'évaluation pour la reconnaissance par la voix organisée chaque année depuis 1997 [Fis07].
- Le *Biometric Test Program* élaboré en 2001 par le Biometrics Working Group au Royaume Uni qui compare six modalités différentes à savoir le visage, l'empreinte digitale, la forme de la main, l'iris, les veines de la main et la voix dans le cadre des systèmes de vérification [MKCK01].
- Le *Face Recognition Vendor Test* (FRVT) en 2000, 2002 et 2006 pour l'évaluation de la reconnaissance du visage 2D/3D, sous une luminosité contrôlée et non contrôlée, et avec des images et des séquences vidéo [PGM⁺03].
- l'Iris Challenge Evaluation (ICE) organisée en 2005 et conjointement avec FRVT en 2006, qui a permis d'évaluer les performances de reconnaissance à partir des images de l'iris des deux yeux [Phi07].

Toutes ces campagnes d'évaluation se sont focalisées sur les systèmes d'authentification et de vérification biométriques. Seule la campagne FRVT 2002 [PGM⁺03, GMP03] de reconnaissance par le visage a conduit à des résultats d'identification de personnes inconnues dans une grande base de données faciales. La complexité d'une telle tâche réside dans la nécessité de calcul d'une matrice de similarité entre toutes les paires de signatures de la galerie et de l'ensemble de test. Dans FRVT 2002, les performances de reconnaissance ont été mesurées sur des bases de données allant jusqu'à 37437 individus. Ainsi, cette campagne d'évaluation a donné la première caractérisation de la manière dont la taille de la base de données affecte les résultats de reconnaissance. Les résultats d'identification des différentes compagnies sur la galerie de 37 437 individus sont présentés sur les figures 1.8 et 1.9.

Les résultats d'identification illustrés sur la figure 1.8 sont des courbes de scores cumulés avec une échelle de rang logarithmique. Comme nous l'avons vu dans le paragraphe 1.5.2.1, une courbe de scores cumulés représente pour chaque rang n des abscisses la probabilité que la réponse désirée se trouve parmi ces n plus proches réponses retournées par le système. Ces différentes courbes de scores cumulés sont calculées sans normalisation de score. En effet, d'autres expérimentations de la même campagne FRVT 2002 sur l'effet de la normalisation des scores de similarités ont montré une amélioration stable des résultats pour les systèmes d'authentification. Cependant, cette normalisation des scores de similarité n'a pas amélioré les performances de l'identification. D'après la figure 1.8, on remarque que même les

meilleurs systèmes atteignent difficilement les 90% de reconnaissance à travers 1000 identités et on voit que le taux d'identification atteint à peine 75% de reconnaissance.

D'autres expérimentations, de la même campagne FRVT 2002, ont examiné la façon dont évolue le taux d'identification en fonction de la taille de la galerie. La figure 1.9 extraite du rapport [PGM⁺03] montre les performances en fonction de la taille de la galerie. L'axe des abscisses est la taille de la galerie à l'échelle logarithmique et l'axe des ordonnées représente le taux d'identification. Les performances ont été calculées pour treize différentes tailles de la galerie : 25, 50, 100, 200, 400, 800, 1600, 3000, 3200, 6400, 12800, 25600, 37437. Les taux d'identification communiqué sont des moyennes sur plusieurs ensembles de galeries. Pour chaque taille de galerie, plusieurs ensembles ont été générés et la moyenne du taux d'identification a été calculée pour donner une bonne estimation de performance.

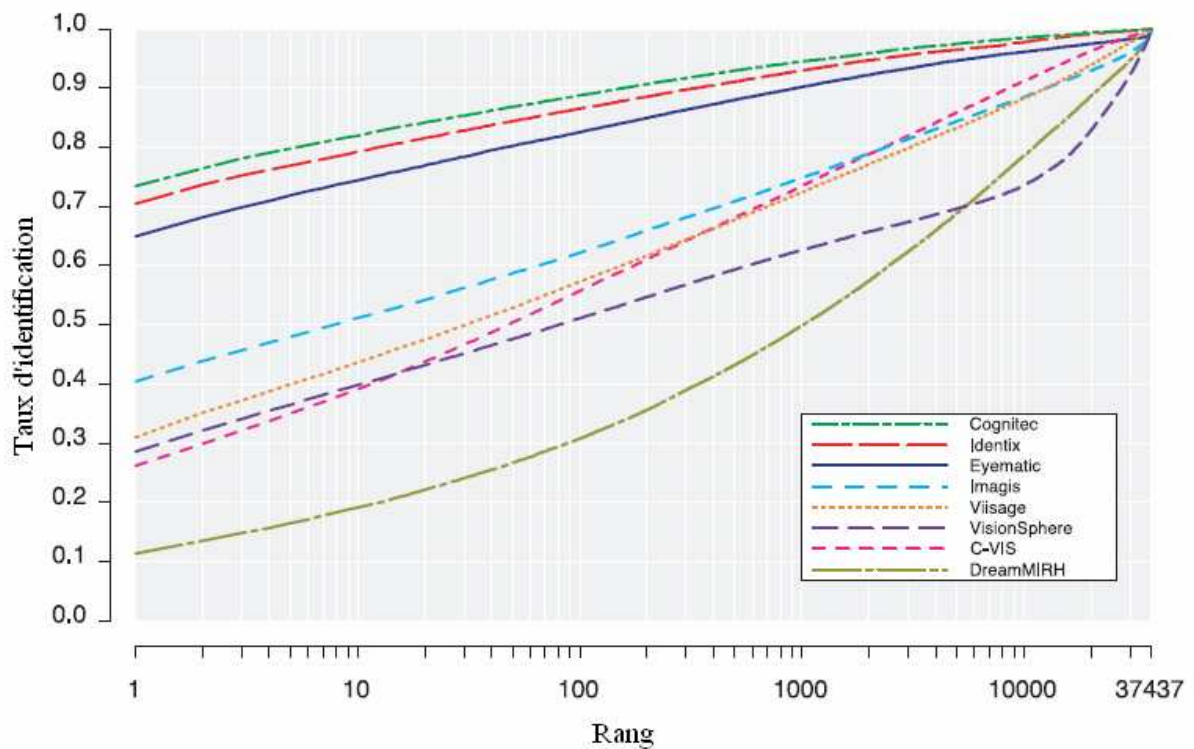


Figure 1.8: Courbes de scores cumulés, communiqués par « FRVT 2002 Evaluation Report » [PGM⁺03], pour une très grande base de données de 37437 personnes.

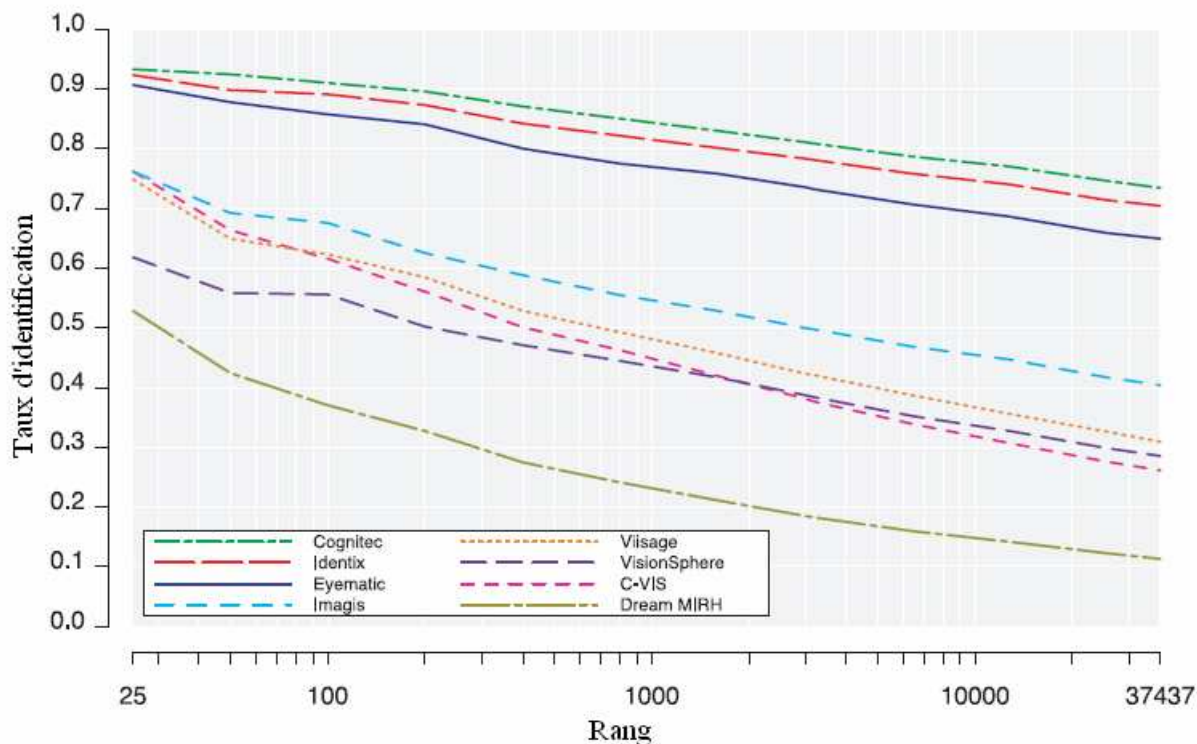


Figure 1.9 : Taux d'identification en fonction de la taille de la galerie, communiqués par « FRVT 2002 Evaluation Report » [PGM⁺03].

1.6.2 Limitations et solution proposée

Il y a trois raisons primaires qui expliquent les performances imparfaites d'un système de reconnaissance biométrique en général [JPP⁺04]. Tout d'abord, la limitation de l'information nécessaire et discriminante pour une modalité biométrique donnée. Ensuite, la limitation du modèle de représentation d'une part, et l'insuffisance des échantillons biométriques disponibles pour extraire des vecteurs caractéristiques qui soient discriminants entre personnes différentes et stables et invariants pour une même personne, d'autre part. Enfin, étant donné la représentation utilisée, la conception d'un module d'appariement qui doit prendre en compte les relations des différents attributs caractérisant un individu reste un problème difficile.

Ces insuffisances posent des limites pour une technologie d'authentification biométrique employant une modalité biométrique, une méthode de représentation des caractéristiques, et un module d'appariement donné. En outre de ces limitations intrinsèques à toutes les technologies de reconnaissance biométriques, les systèmes d'identification dans les grandes bases de données biométriques connaissent une chute de performances. Cela est dû au facteur d'échelle combiné à ces dernières limitations intrinsèques d'une méthode de reconnaissance biométrique.

Ainsi, nous remarquons sur la figure 1.9 que les trois meilleurs systèmes évalués dans la campagne FRVT 2002 connaissent une baisse de 20 à 25% de leurs taux d'identification en passant d'une galerie de petite taille à une galerie de plusieurs milliers d'utilisateurs.

Il y a deux approches par lesquelles nous pouvons essayer de réduire l'erreur de tels systèmes d'identification :

- (i) par la réduction des erreurs intrinsèques de la méthode de reconnaissance à savoir les *FAR* et *FRR* correspondant,
- (ii) par la réduction de l'espace de recherche en amont de l'identification.

Le *FAR* et le *FRR* d'une modalité biométrique donnée sont limités par les différentes variabilités affectant la modalité biométrique utilisée. Aussi, ils sont limités par l'algorithme d'identification utilisé et ne peuvent pas être réduit indéfiniment du fait des raisons limitatives, d'une méthode de reconnaissance, expliquées au début de cette section. Le tableau 1.3 présente, à titre indicatif, les performances intrinsèques de différentes modalités biométriques. Les résultats calculés pour des *FRR* de l'ordre de 1%, sont extraits de [JPP⁺04].

Biométrie	Visage	Empreinte digitale	Forme de la main	Iris
FAR	10%	0,01%	1,5%	0,001%

Tableau 1.3 : Performance typique de chaque modalité biométrique évaluée pour des *FRR* de l'ordre de 1% [JPP⁺04].

L'approche qui s'impose pour améliorer la vitesse et la précision d'un système biométrique d'identification est la réduction du nombre de personnes de la galerie avec lesquelles l'appariement est effectué à la suite d'une requête. Ceci exige que les caractéristiques et les attributs des personnes enrôlées dans la base de données soient classifiés, divisés et indexés d'une manière discriminante.

Les systèmes de réduction de l'espace de recherche ont été essentiellement étudiés à travers les systèmes d'identification des empreintes digitales [MMJP03]. Nous étudions, dans ce qui suit, les effets de réduire l'espace de recherche, appelé aussi galerie ou base de données du système, pendant l'identification.

Supposons que nous pouvons ramener l'espace de recherche à une fraction *R* de la base de données entière. Le *FAR* et le *FRR* sont alors donnés par :

$$FAR_{RN} = 1 - (1 - FAR)^{R \cdot N} \approx (R \cdot N) \cdot FAR \quad (1.11)$$

où *R*, appelé *taux de pénétration*, est le pourcentage de la galerie retenu pour la suite de la recherche : $0 \leq R \leq 1$.

$$FRR_{RN} = RER + (1 - RER) \cdot FRR \quad (1.12)$$

où *RER*, le taux d'erreur de la recherche (*Retrieval Error Rate*), est la probabilité que la signature correspondante à l'identité recherchée soit rejetée de la galerie par le mécanisme de partitionnement et de réduction de la base de données.

Nous remarquons que plus le taux de pénétration *R* est petit, plus on réduit la base de données du système (galerie) pour la suite de l'identification. Le taux de fausse acceptation *FAR_{RN}* du système de reconnaissance correspondant est ainsi réduit. Cependant, le taux de faux rejet *FRR_{RN}* a tendance à augmenter à cause de l'apparition d'un taux d'erreur de la recherche *RER*

correspondant au rejet de l'identité recherché de la partition retenue pour la suite de l'identification.

Nous proposons, dans cette thèse, de réduire au maximum la taille de la galerie sans compromettre les performances d'un système d'identification qui peut utiliser une technique d'identification différente et plus élaborée que celle utilisée pour la partition de la base de données (galerie).

La multi représentation (représentation multiple d'une modalité biométrique par différentes signatures) peut être d'un grand apport dans le cadre d'une architecture générique de tels systèmes d'identification. Nous proposons dans le chapitre V cette architecture composée d'un module de classification suivi par un module d'identification. En effet, la réduction de l'espace de recherche (galerie) consiste à classer ou à catégoriser la base de données initiale sur plusieurs sous ensembles. L'utilisation intelligente de méthodes de représentation différentes pour cette phase de classification et pour la phase finale d'identification sur l'espace réduit, doit accroître les capacités de reconnaissance du système. Nous cherchons dans le cadre de cette thèse, parmi les quelques méthodes que nous avons pu évaluer de la reconnaissance faciale, les méthodes de représentation et de classification optimales pour cette finalité.

Ainsi, en limitant l'identification finale à seulement une fraction de la base de données entière, nous pouvons assurer, non seulement une diminution du temps de réponse, mais également une amélioration de la précision du système global.

D'autre part, cette solution reste valable en monde ouvert aussi bien qu'en monde fermé. En effet, on ne risque pas un faux rejet lorsque l'utilisateur n'est pas enregistré dans la base de données du système. Dans ce cas, plus on réduit la galerie par le module de classification, plus le système sera capable de juger que l'utilisateur est inconnu par rapport à une base d'individus réduite.

1.7 Etat de l'art de systèmes de classification des signatures biométriques

1.7.1 Systèmes de classification des empreintes digitales

Plusieurs travaux ont été établis quant à la classification des empreintes digitales suivant leurs caractéristiques morphologiques globales [NB97, JP00, TBL03, GZ07, FJ08]. Les variations d'orientation des stries génèrent des points singuliers globaux (les centres et les deltas). Leurs quantités et leurs positions relatives permettent de classer les empreintes digitales généralement en cinq classes (cf. figure 1.10) :

- Les arches ;
- Les deltas ;
- Les boucles à droite ;
- Les boucles à gauche ;
- Les spires.

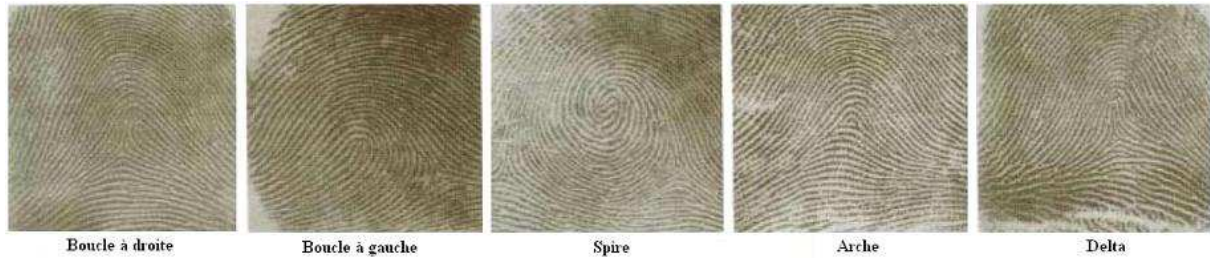


Figure 1.10 : Les cinq classes des empreintes digitales.

Cependant, les empreintes digitales ne sont pas uniformément distribuées parmi ces cinq classes : les proportions sont 3,7%, 2,9%, 33,8%, 31,7% et 27,9% pour les arches, les deltas, les boucles gauches, les boucles droites et les spires, respectivement [CMMN03]. Ainsi, trois de ces cinq classes représentent plus de 93% de la population totale.

1.7.2 Classification des images faciales suivant le genre et/ou la race ethnique

Plusieurs travaux ont été développés pour la détermination du genre ou du sexe humain à travers les images du visage. Différentes approches pour résoudre ce problème ont été proposées dans la littérature. Ces techniques incluent les techniques basées sur l'apparence telles que l'analyse en composantes principales et les ondelettes de Haar [VAET97, LCJ06, XLS08], les méthodes basées sur les caractéristiques géométriques [BBD93, WS08], les méthodes d'appariement de modèles [BP93], les méthodes d'appariement de graphes [WFKM95], et les méthodes basées sur les réseaux de neurones [HB95, MR08]. Tous ces travaux rapportent un pourcentage de classification de l'ordre de 85 à 95% pour des bases d'images différentes ne dépassant pas les 100 images. De plus récents travaux utilisant les réseaux de neurones LVQ (Learning Vector Quantization) et les réseaux RBF (Radial Basis Function) [Tol01], les Séparateurs à Vastes Marges (SVM) [MY00, CBA04, AG05, LL06, YLA06] ainsi que d'autres algorithmes et stratégies de classification comme le dopage (boosting) adaptatif [SSNT06, BR07] rapportent de meilleurs résultats. Des bases d'évaluation plus conséquentes de plusieurs centaines d'individus ont été utilisées pour tester ces systèmes avec notamment la base FERET [PWHR98]. Les meilleurs taux atteignent 94% pour l'identification du sexe et 96% pour la classification ethnique en deux catégories : asiatique et non asiatique.

Cependant, les taux d'erreurs induits par ces classifieurs (autour de 6% et de 4% respectivement pour discriminer le genre et la catégorie ethnique) sont trop élevés pour espérer une réduction efficace de la galerie et pour utiliser ce processus dans le cadre d'un système d'identification biométrique. En effet, les taux d'erreurs résultant de ces classifications sont équivalents au taux d'erreur de la recherche *RER*. L'équation 1.12 démontre l'effet de ce dernier taux d'erreur sur le rejet des identités recherchées et par conséquent sur la précision de l'identification. Le taux d'erreur de la recherche *RER* peut ainsi pénaliser davantage la précision du système si on procède à une réduction de la galerie sur la base de la partition élaborée. Ainsi, un grand challenge reste à relever quant à la classification des visages avec les différentes contraintes liées principalement aux conditions d'éclairages et aux expressions faciales.

1.7.3 Classification non supervisée pour la reconnaissance biométrique

Peu de travaux ont cherché à classifier sans a priori (*clustering*) des modalités biométriques pour lesquelles on n'a pas d'évidentes caractéristiques morphologiques globales définissant des catégories apparentes comme pour les empreintes digitales.

Les auteurs dans [MPCG05, MCG05] ont proposé une approche qui effectue une classification grossière basée sur les attributs de la signature et de la forme de la main avant de procéder à l'appariement exhaustif. La base de données est divisée en plusieurs classes utilisant l'algorithme de classification k-moyennes. Etant donné une personne à identifier, ses caractéristiques sont assignées à une certaine partie de la base de données et toutes les identités de cette partition sont considérées comme des candidats potentiels pour l'étape finale d'identification.

En outre, de telles méthodes de classification d'échantillons biométriques n'existent pas pour d'autres modalités comme le visage et l'iris. Une méthodologie généralisée plus sophistiquée pour classifier les caractéristiques biométriques est nécessaire pour réaliser une recherche efficace dans les bases de données biométriques. Ainsi, l'utilisation de plusieurs modalités biométriques indépendantes ne présentant pas de corrélation entre leurs caractéristiques peut augmenter les performances du système d'identification [All09].

1.8 Conclusion

Nous abordons dans cette thèse le problème de la classification automatique non supervisée de caractéristiques biométriques sans aucun a priori sur la nature des classes élaborées. Notre objectif est de regrouper les caractéristiques biométriques au sens d'un critère donné pour simplifier par la suite la galerie de recherche d'identité et développer l'identification sur un sous espace d'identités réduit.

Une position générale de la problématique d'identification dans les grandes bases de données biométriques peut être formulée comme suit : étant donné une personne à identifier, ses caractéristiques biométriques sont extraites et comparées à des partitions déjà élaborées de la galerie. Cette technique sera employée pour rétrécir la recherche, augmentant ainsi la précision tout en diminuant la complexité du problème d'identification. La solution automatique du problème consiste à effectuer des classifications des différentes caractéristiques acquises de la personne en question. Les étapes de la classification automatique et de l'identification finale peuvent s'appuyer sur une combinaison des différentes caractéristiques à savoir les visages 2D/3D, les empreintes digitales, l'iris, la forme de la main, ... Ainsi une méthodologie généralisée plus sophistiquée pour classifier les caractéristiques biométriques est nécessaire pour réaliser une recherche efficace dans les grandes bases de données biométriques.

Nous nous limitons dans le cadre de cette thèse à l'étude de différentes techniques de classification et d'identification sur les caractéristiques faciales. Cependant le schéma conceptuel de notre proposition détaillé dans les chapitres 5 et 6 reste valable pour toute autre modalité biométrique. L'utilisation de techniques différentes et indépendantes de reconnaissance permettront d'améliorer encore plus les performances de notre système en augmentant la précision tout en décroissant la complexité de l'identification.

Nous présentons dans le chapitre suivant, un travail de localisation de visage, effectué en collaboration avec Mohamed Saaidia. La localisation du visage est le premier maillon de la chaîne de la reconnaissance faciale.

Chapitre II

Détection et localisation de visages - Vers la reconnaissance automatique des visages

2.1 Introduction

Dans le spectre de la biométrie, la reconnaissance de visages revêt un caractère particulier du fait du grand nombre de travaux qui lui sont consacrés [FRH.org]. En effet capturer une image d'un visage, en particulier à travers une caméra 2D, est simple et non invasif. C'est donc une modalité biométrique facilement tolérée par les utilisateurs, et ceci d'autant plus que les ordinateurs sont équipés de plus en plus souvent de « webcam ». Mais les performances de la reconnaissance faciale sont toujours bien en deçà de ce que l'on pourrait espérer pour de telles applications. L'augmentation des taux de reconnaissance, la diminution des reconnaissances à tort et l'accélération des temps de réponse lors de la recherche dans les grandes bases de données biométriques sont les défis auxquels doivent faire face les algorithmes actuellement développés.

Pour reconnaître une personne à partir de son image, il faut passer par certaines étapes. Tout d'abord il faut détecter et/ou localiser le visage dans l'image. Ensuite il faut éventuellement le normaliser pour ramener le visage à une taille standard. Puis il s'agit d'aborder la phase de reconnaissance proprement dite. Dans ce chapitre nous proposons de développer deux des étapes clés de cette chaîne de reconnaissance automatique à savoir la localisation et la normalisation des visages. Nous développons ainsi deux approches concurrentes pour localiser un visage, l'une basée sur les moments de Zernike et l'autre sur les *Eigenfaces*. Nous détaillons ensuite notre phase de normalisation pour centrer et aligner les visages. Les images utilisées pour les différents résultats présentés sont issues de la base XM2VTS [Mes99].

2.2 Chaîne de traitement préalable à la reconnaissance automatique

La détection d'un visage dans une image ou dans une séquence d'images revêt aujourd'hui une importance capitale dans plusieurs domaines d'applications qui traitent des aspects liés à la communication Homme-Machine, à la compression de l'information, à l'estimation de pose etc. Dans le cadre de la reconnaissance automatique du visage, certaines techniques reposent sur la mise en œuvre simultanée de la détection des visages et la détection des caractéristiques faciales. Cette dernière tâche ouvre aussi un large spectre d'applications connexes telles que l'analyse d'émotion et l'estimation de la direction du regard. Nous illustrons ci-dessous le processus de prétraitements qui vise à segmenter et à extraire l'image du visage. Ce processus se décompose généralement 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'est-à-dire 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.

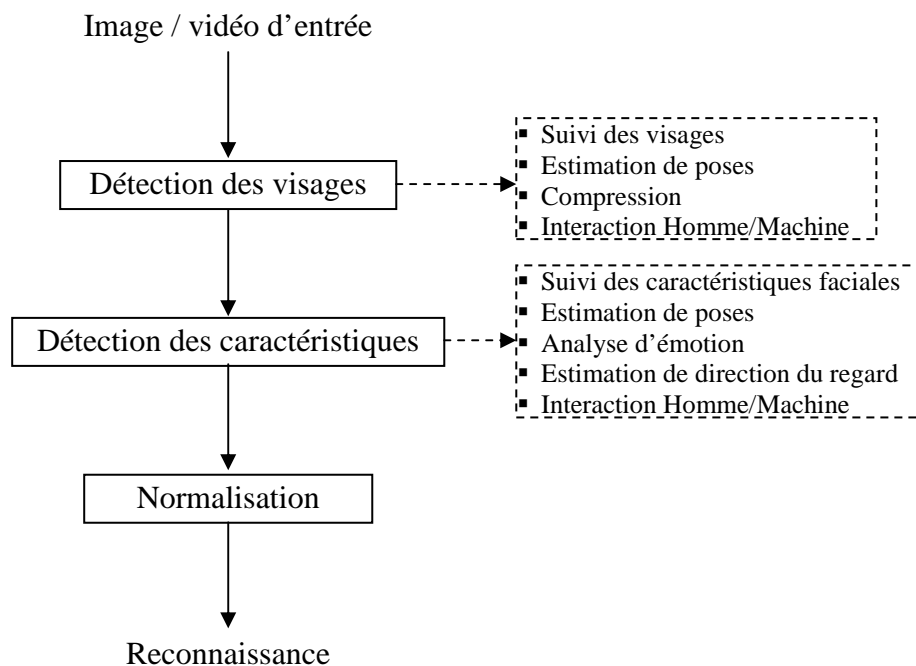


Figure 2.1 : Prétraitements des images en amont de la reconnaissance. Les différentes étapes du prétraitement peuvent également servir pour d'autres applications (encadrés en pointillés).

A la différence de ces approches qui analysent le visage dans sa globalité pour arriver aux caractéristiques locales, d'autres méthodes commencent par détecter certaines caractéristiques du visage pour localiser ensuite le visage entier dans une image [YKA02].

Détecter et reconnaître une personne nécessite un apprentissage de ses caractéristiques faciales inhérentes et remarquables. Partant du même principe de la reconnaissance humaine des visages, la détection et la reconnaissance automatique nécessite un processus d'apprentissage qui diffère selon les modèles et les techniques mis en œuvre. Nous exposons dans le paragraphe suivant les différentes formes d'apprentissage automatiques qui peuvent être induit à la fois par la détection et la reconnaissance faciale.

2.3 La détection et la reconnaissance faciale - un processus d'apprentissage

La perception, en général, et la reconnaissance faciale en particulier, ont besoin d'apprentissage. C'est à travers l'immersion dans notre monde et à travers notre développement que nous apprenons graduellement à voir et à reconnaître les objets, et en particulier les visages humains [Wec06].

Nous nous plaçons dans le domaine de la classification non supervisée (*clustering*) ou apprentissage sans professeur. Plusieurs techniques non supervisées ont été développées pour répondre à ces problèmes. Il s'agit d'un ensemble de techniques qui s'inspirent de processus d'apprentissage mis en œuvre par un être humain et visant à doter la machine de cette capacité.

L'apprentissage est essentiellement relatif à la sélection du modèle et à la prédiction. Selon Simon [Sim82], l'apprentissage signifie : « les changements adaptatifs qui permettent au système de faire plus efficacement la même tâche dans le futur ». En particulier,

l'apprentissage modélise les associations qui sont alors utilisées pour inférer des sorties étant donné les entrées correspondantes, c'est-à-dire, la localisation et l'identification à partir d'une représentation faciale donnée dans notre cas d'application. L'apparence faciale varie énormément bien que son identité reste constante. Le degré d'une détection et d'une reconnaissance faciale fiable est la capacité de généralisation quant aux différentes formes et aux différentes variabilités qui peuvent l'affecter. Comme la généralisation est désirée pour la plupart des systèmes de vision par ordinateur, la seule solution est d'y incorporer la capacité d'apprentissage.

Un processus d'apprentissage prédictif peut être réalisé de manière déductive ou de manière inductive [CM98]. La différence essentielle entre l'apprentissage déductif et l'apprentissage inductif est la façon de construire les règles. En apprentissage déductif, des nouvelles règles sont déduites à partir des anciennes. En apprentissage inductif (apprentissage par l'exemple), on essaie de trouver des règles à partir d'un ensemble de cas connus, structurés dans une base d'apprentissage, selon une méthode de généralisation, c'est-à-dire du particulier au général. Chaque cas, appelé « exemple » ou « individu », est décrit par un ensemble d'attributs et une classe qui lui est associée. Le but ici est d'apprendre l'association entre les attributs et les classes. Dans ce cadre d'apprentissage inductif, obtenir un bon modèle d'inférence nécessite [Wec06] :

- la spécification d'un modèle (ou fonction d'approximation) ;
- un principe inductif pour combiner le modèle avec les données ;
- et une procédure d'optimisation (apprentissage) pour estimer les paramètres du modèle.

Après avoir obtenu des règles qui résument ces exemples, le système peut traiter de nouveaux cas et déterminer leur classe. Les règles doivent donc être capables de discriminer les exemples des différentes classes. Une troisième forme d'apprentissage transductif a été introduite récemment par Vapnik [Vap98, Vap00]. L'apprentissage transductif implique une forme d'inférence locale qui estime à partir des exemples, non pas les règles (ou les fonctions d'approximations) comme l'apprentissage inductif, mais les valeurs de ces fonctions à des points d'intérêts.

L'apprentissage inductif est l'une des approches utilisées pour résoudre des problèmes d'extraction de connaissances à partir de données (comme l'Analyse en Composante Principale largement utilisée dans la détection et la reconnaissance faciale) et de classification. Il apparaît, entre autre, dans plusieurs types d'applications, par exemple, en classification, en prédiction, en extraction de règles à partir de données, en résumé de données, ou encore en généralisation [Dan07]. Plusieurs techniques d'apprentissage ont été proposées et utilisées, nous pouvons citer comme exemples : les réseaux neuronaux, les arbres de décision, les réseaux bayésiens, ... Parmi ces techniques, nous nous intéressons, dans le cadre de cette thèse, aux réseaux de neurones pour la localisation faciale et à différentes techniques d'apprentissage non supervisé pour la classification automatique des visages en vue de simplifier le processus d'identification.

2.4 Détection et localisation de visage - Etat de l'art

La détection de visage a pour objectifs de décider si une image contient un ou plusieurs visages, et de donner leur localisation. Deux principales métriques caractérisent les performances d'un système de détection : le taux de détection, c'est-à-dire, le pourcentage de visages correctement détectés, et le taux de faux positifs, correspondant à des détections dans

des régions ne contenant pas de visage. Un système performant est associé à la fois à un taux de détection important et à un taux de faux positifs faible. D'autre part, la localisation faciale, consiste à trouver la position exacte du visage dans une image. La mesure de la qualité d'un algorithme de localisation demeure encore un point sensible à résoudre et auquel Saaidia [SLVB07] a proposé une mesure objective que nous explicitons dans le paragraphe 2.5.4.

Les premiers travaux portant sur la détection de visages dans des images à fond simple ou complexe remontent au milieu des années 1990. Deux états de l'art détaillés et relativement récents sont fournis dans [Hje01, YKA02]. Les approches de détection peuvent être divisées en deux catégories : les approches *locales* et les approches *globales*.

2.4.1 Les approches locales

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 [YKA02, GS06]. D'autres approches utilisent la mise en correspondance de modèles locaux, statistiques ou déformables [AB05, WY06]. Ces modèles sont généralement localisés autour des caractéristiques faciales (yeux, nez, ...) 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 à fournir au système, par exemple, une certaine robustesse aux occultations partielles des visages. La recherche des caractéristiques faciales étant menée dans l'intégralité de l'image, elle est fortement influencée par le bruit de l'image. Par conséquent, l'imprécision en termes de détection de caractéristiques faciales et le nombre de faux positifs sont potentiellement importants, ce qui dégrade les performances de ces techniques de détection de visages.

2.4.2 Les approches globales

Les méthodes globales, c'est-à-dire 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 grande variabilité dans les conditions d'éclairage, 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. La majorité de ces approches 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 connues reposent sur l'utilisation de techniques d'analyse statistique multivariées [CH97, SP98, SK00, GZT00, YAK01, Liu03] pour la représentation des visages et/ou leur classification. Cette phase d'analyse est éventuellement suivie 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 pour la classification et la détection de visages [OFG97, RBK98, FBVC01, VJ01, WBM08].

2.5 Notre approche de localisation

L'objectif de ce module est de localiser au mieux le visage pour espérer optimiser la reconnaissance faciale à suivre. Nous opérons ainsi dans un contexte biométrique avec des

visages sur un fond relativement neutre mais avec des poses, des éclairages et des expressions différents. La méthode proposée ici exploite des caractéristiques globales du visage pour inférer sa localisation précise dans l'image. Nous avons collaboré avec Mohamed Saaidia qui a effectué une partie de ses travaux de thèse sur la détection et la reconnaissance d'expressions faciales dans notre équipe au sein du laboratoire IBISC. Dans [SCL⁺07, CLS⁺08], nous avons proposé l'utilisation des moments de Zernike comparés aux caractéristiques *Eigenfaces* pour une localisation fine et adaptée du contour d'un visage. En effet, les moments de Zernike ont été récemment développés pour la reconnaissance de visage [HAF03, Ali05], mais n'ont jamais été utilisés pour la détection générale d'objets et de visages en particulier. Ainsi, Haddadnia [HAF03] et Alirezadee [Ali05] ont utilisé, dans leurs travaux sur la reconnaissance des visages, une méthode de localisation du visage basée région qui consiste en une segmentation par détection de contours suivie d'une localisation d'éléments connexes dont l'approximation par une ellipse (représentant la forme générale du visage) et un seuillage déterminent si on est en présence ou non d'un visage. Ono [Ono03] dans un travail similaire, a utilisé une méthode basée sur la localisation de l'espace entre sourcils à l'aide d'un filtre à fréquence de résonance. Dans des travaux concernant l'analyse d'images contenant des objets, autres que des visages, par les moments de Zernike la même constatation est faite. Rosenberger et al [RRE04], par exemple, utilisent une procédure de segmentation pour la localisation d'objets qu'ils décrivent ensuite par moments de Zernike.

Cependant, la définition et la formulation des moments de Zernike comme étant des paramètres capables de traduire les informations géométriques d'une fonction bidimensionnelle et de les compresser sous forme d'un vecteur de longueur suffisamment réduite permettent de prétendre à leur utilisation dans un objectif de détection d'objets dans une image. En effet, les moments géométriques ne sont pas des paramètres abstraits mais plutôt des paramètres explicites ou explicitables suivant une signification géométrique de l'image. Chaque moment de l'image représente une caractérisation géométrique bien définie tels que la surface, le centre de masse, les symétries horizontale et verticale... Ainsi, un visage par sa forme particulière et son contenu géométriquement riche par les détails des éléments qu'il contient (yeux, bouche, sourcils, ...) pourrait avoir une certaine prépondérance au niveau du vecteur Zernike représentant une image le contenant.

Nous avons proposé, d'autre part, une étude comparative de la qualité de la localisation entre les moments de Zernike et les *Eigenfaces*. En effet, les *Eigenfaces* furent le premier type de caractérisation utilisé avec succès dans la détection et la reconnaissance faciale [TP91]. Cette méthode est basée sur la décomposition de l'image suivant les axes de variations prépondérants d'un ensemble d'apprentissage afin d'extraire un vecteur caractéristique de taille réduite représentant l'image faciale.

Le problème de détection du visage sera considéré et traité comme un problème de classification. On se retrouve alors devant les trois difficultés classiques de ce genre de problèmes, à savoir, la non linéarité, la dimensionnalité importante de l'espace de représentation et la nécessité d'aboutir à une implémentation généralisée. Les réseaux de neurones paraissent alors comme une alternative intéressante pour notre problème. Ces constatations nous ont donc encouragées à mettre en œuvre une méthode de localisation du visage en utilisant les moments géométriques de Zernike et les vecteurs *Eigenfaces* comme entrées d'un réseau de neurones. Ce réseau de neurone, préalablement entraîné sur la base de vecteurs caractéristiques, délivre à sa sortie un ensemble de points représentant un contour probable du visage recherché.

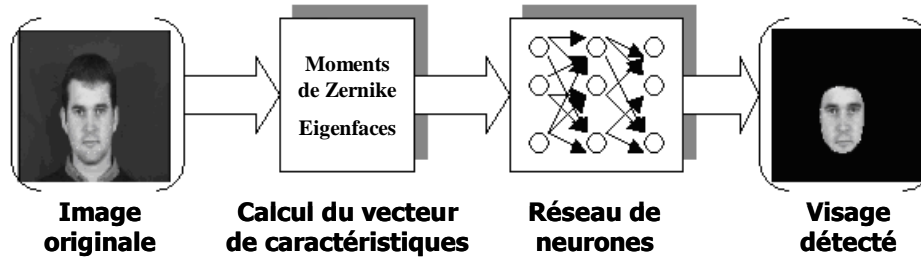


Figure 2.2 : Schéma de principe de notre méthode de localisation du visage.

Les deux méthodes de localisation de visage que nous avons développées relèvent de la catégorie des méthodes globales. Elles diffèrent seulement dans leur première phase de mise en œuvre, comme l'illustre la figure 2.2, c'est-à-dire par la technique utilisée pour caractériser l'image à traiter. La première méthode utilise les propriétés géométriques à travers le calcul des moments de Zernike alors que la seconde est basée sur une analyse multifactorielle pour extraire un sous espace de vecteurs propres : *Eigenfaces*. La deuxième étape de ce processus de détection est assurée par un réseau de neurones MLP de type *back-propagation* entraîné par les vecteurs caractéristiques calculés à partir des *Eigenfaces* ou des moments de Zernike. Le réseau de neurones fournit comme sortie un ensemble de points représentant le contour probable du visage.

Pour permettre une évaluation objective des résultats de localisation de chacune des deux approches, un critère de qualité est proposé dans le paragraphe 2.5.4 permettant de définir un indicateur de comparaison quantitatif des résultats obtenus.

2.5.1 Les moments de Zernike

Les moments de Zernike ont été introduits par F. Zernike en 1934 [Zer34]. Dans le domaine du traitement de l'information, les moments de Zernike ont beaucoup été utilisés pour leur propriété d'orthogonalité qui permet la génération de descripteurs non redondants et leurs propriétés d'invariance en translation, en échelle et en rotation. Ainsi, on retrouve les moments de Zernike dans beaucoup de travaux concernant la reconnaissance d'images de personnes [RY98], l'indexation d'images dans les bases de données, l'analyse et la description de forme d'objet 2D ou 3D...

La formulation de ces moments est donnée par :

$$Z_{nm} = \frac{n+1}{\pi} \int \int_{x,y} f(x,y) \cdot [V_{nm}(x,y)]^* dx dy \quad \text{où } x^2 + y^2 \leq 1 \quad (2.1)$$

[...]^{*} est utilisé pour indiquer la valeur complexe conjuguée.

n représente ici l'ordre de décomposition ($n=0,1,2,\dots,\infty$), dit aussi ordre radial, et m le nombre de répétitions de la décomposition ou fréquence azimutale pour un ordre n donné. L'ordre et la répétition sont liés par les deux conditions suivantes :

$$n - |m| \text{ toujours pair} \quad \text{et} \quad |m| \leq n \quad (2.2)$$

$V_{nm}(x,y)$ représente les polynômes de Zernike constituant la base orthogonale de projection. Ils s'écrivent en général en représentation polaire sous la forme suivante :

$$V_{nm}(r, \theta) = R_{nm}(r) \cdot \exp(jm\theta) \quad (2.3)$$

Où $R_{nm}(r)$ sont des polynômes radiaux de la forme :

$$R_{n,m}(r) = \sum_{k=|m|}^n \frac{(-1)^{(n-k)/2} \cdot (n+k)!}{\left(\frac{n-k}{2}\right)! \left(\frac{k+m}{2}\right)! \left(\frac{k-m}{2}\right)!} r^k \quad (2.4)$$

L'application des moments de Zernike à une fonction discrète $h(x,y)$ (image par exemple) nécessite la réécriture de (2.1) comme suit :

$$Z_{nm} = \frac{n+1}{\pi} \sum_x \sum_y h(x,y) \cdot [V_{nm}(x,y)]^* \quad (2.5)$$

où $x^2 + y^2 \leq 1$

2.5.1.1 Approximations

Afin de conserver l'orthogonalité de la base au niveau des moments calculés, la fonction $h(x,y)$ doit être recalculée à l'intérieur du cercle unité en la transférant en représentation polaire (r,θ) tel que le centre de l'image soit le centre du cercle unité. Les relations de transformation s'écriront alors :

$$x = r \cdot \cos \theta \quad \text{et} \quad y = r \cdot \sin \theta \quad (2.6)$$

avec $r = \sqrt{x^2 + y^2}$ et $\theta = \tan^{-1}\left(\frac{y}{x}\right)$.

La forme de l'image $h(x,y)$ est rectangulaire ou carré ce qui est incompatible avec la forme du cercle unité. Ceci nous impose de faire un choix entre l'élimination de certains points de l'image (les coins surtout) ou bien l'introduction de points étrangers à la fonction originale.

On réécrit les relations de (2.6) comme suit :

$$\begin{cases} x_j = c + \frac{j \cdot (d-c)}{N-1} \\ y_i = d - \frac{i \cdot (d-c)}{M-1} \end{cases} \quad \text{et} \quad \begin{cases} r_{ij} = \sqrt{x_j^2 + y_i^2} \\ \theta_{ij} = \tan^{-1}\left(\frac{y_i}{x_j}\right) \end{cases} \quad (2.7)$$

Où i et j sont les coordonnées du point de l'image originale. x_j et y_i les nouvelles coordonnées de ce point dans le nouveau repère (le cercle unité). M et N sont respectivement l'étendue horizontale et verticale de cette image. c et d sont les paramètres qui permettent de faire le choix entre recalculer la fonction $h(x,y)$ complètement ($c = -1/\sqrt{2}$ et $d = 1/\sqrt{2}$) ou partiellement ($c = -1$ et $d = 1$) dans le cercle unité comme le montre la figure 2.3.

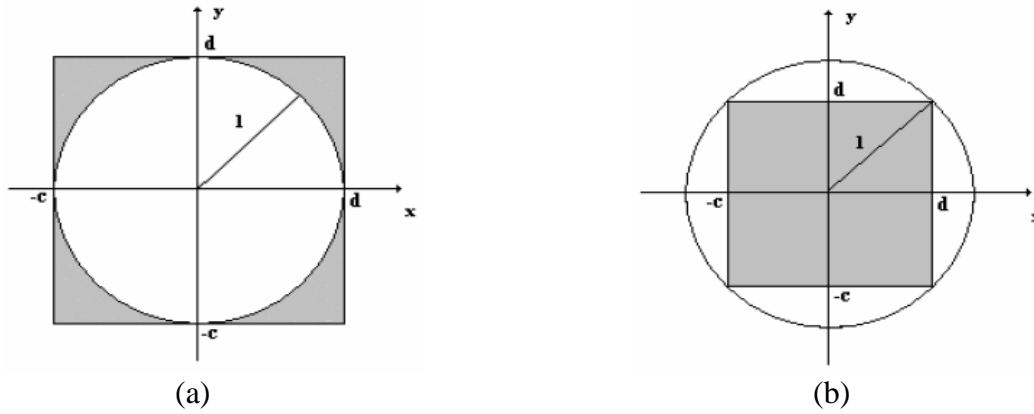


Figure 2.3 : Façon de recalculer de la fonction $h(x,y)$ (rectangle gris) dans le cercle unité :

a) cas de $c = -1$ et $d = 1$; b) cas de $c = -1/\sqrt{2}$ et $d = 1/\sqrt{2}$.

2.5.1.2 Mise en œuvre des moments de Zernike

La formulation classique des moments de Zernike détaillée ci-dessus est très facile à mettre en œuvre algorithmiquement. Cependant, elle reste très coûteuse en termes de complexité et de temps de calcul et se prête mal à un traitement rapide. Ce handicap majeur a poussé les chercheurs à essayer de trouver une formulation plus appropriée du point de vue complexité algorithmique. Mukandan et Ramakrishnan [MR95] ont proposé un algorithme récursif de calcul des moments de Zernike en coordonnées polaires. Belkasim et al. [BAS96] ont utilisé une expansion radiale et angulaire des polynômes orthonormaux de Zernike pour proposer un algorithme récursif rapide. Gu et al. [GSTL02] ont utilisé une transformation circulaire des relations proposées par Mukandan et al. [MR95] pour essayer d'aboutir à un algorithme plus rapide. Finalement, Amayeh et al. [AEBN05] ont proposé un algorithme qui est aussi rapide que les précédents mais qui permet de préserver une précision de calcul au même niveau que la formulation classique et ce en n'introduisant aucune quantification au niveau des paramètres angle θ et rayon r . La préservation de la précision assure le maintien de l'orthogonalité de la base construite. L'algorithme proposé repose sur la détection, dans le processus d'obtention des moments, de termes communs dont le calcul se fait de la même façon pendant toutes les itérations. Ces termes sont alors calculés une seule fois et mémorisés dans une table à laquelle on accède directement au cours de la réalisation des différentes itérations.

C'est cette formulation que nous avons donc adoptée pour notre application. Pour aboutir à cette forme de représentation, les équations précédentes sont alors réécrites et réorganisées comme le montre les équations (2.8).

$$\begin{aligned}
 Z_{n,m} &= \frac{n+1}{\pi} \sum_{x^2+y^2 \leq 1} \sum_{k=|m|}^n \left(\sum_{k=|m|}^n \beta_{n,m,k} \cdot \rho^k \right) e^{-j.m.\theta} \cdot f(x_j, y_i) \\
 &= \frac{n+1}{\pi} \sum_{k=|m|}^n \beta_{n,m,k} \cdot \left(\sum_{x^2+y^2 \leq 1} e^{-j.m.\theta} \cdot \rho^k \cdot f(x_j, y_i) \right) \\
 &= \frac{n+1}{\pi} \sum_{k=|m|}^n \beta_{n,m,k} \cdot \mathbf{X}_{m,k}
 \end{aligned} \tag{2.8}$$

où : $\beta_{n,m,k} = \frac{(-1)^{(n-k)/2} \cdot (n+k)!}{\left(\frac{n-k}{2}\right)! \left(\frac{k+m}{2}\right)! \left(\frac{k-m}{2}\right)!}$ est un terme dont le calcul ne dépend ni de l'image

$f(x_j, y_i)$ ni des coordonnées (x_j, y_i) et $X_{m,k}$ est le terme commun qu'on ne calcule qu'une seule fois pour toutes les répétitions. L'équation (2.8) réduit donc le calcul des moments de Zernike de toute image au calcul d'une combinaison linéaire de ces deux derniers termes.

2.5.2 Introduction aux « Eigenfaces »

Les *Eigenfaces* furent le premier type de caractérisation utilisé avec succès dans des traitements faciaux tels que la détection et la reconnaissance du visage. Cette méthode est basée sur la décomposition de l'image traitée selon plusieurs axes de variation par rapport à une image moyenne.

Utilisant initialement les projections de Karhunen Loeve (*KL*) ou de l'Analyse en Composante Principale (*PCA*) pour la reconstruction des visages en dimension réduite [KS90, SK87], les « *Eigenpictures* » ont été une des forces principales pour la représentation et par suite pour l'identification des visages. Il est bien connu qu'il existe des redondances statistiques significatives dans les différents types d'images [Rud94]. Pour une classe limitée des objets telles que les images de visage qui sont normalisées aux facteurs d'échelle, à la translation, et à la rotation, la redondance est encore plus grande [PA96; Zha03]. Une des meilleures représentations compactes globales est l'Analyse en Composantes Principales *PCA*, qui décorrèle les vecteurs de données.

Chaque image I_i , de dimension $(n*m)$, de l'ensemble d'apprentissage est représentée comme un vecteur Γ_i . Soit $\{\Gamma_1, \Gamma_2, \dots, \Gamma_N\}$ l'ensemble des vecteurs images de la base d'apprentissage. Le visage moyen de cet ensemble est défini par :

$$\psi = \frac{1}{N} \sum_{i=1}^N \Gamma_i \quad (2.9)$$

Chaque visage est ensuite centré par le vecteur $A_i = \Gamma_i - \psi$. Un exemple de l'ensemble d'apprentissage est présenté dans la figure 2.4a, avec le visage moyen dans la figure 2.4b.

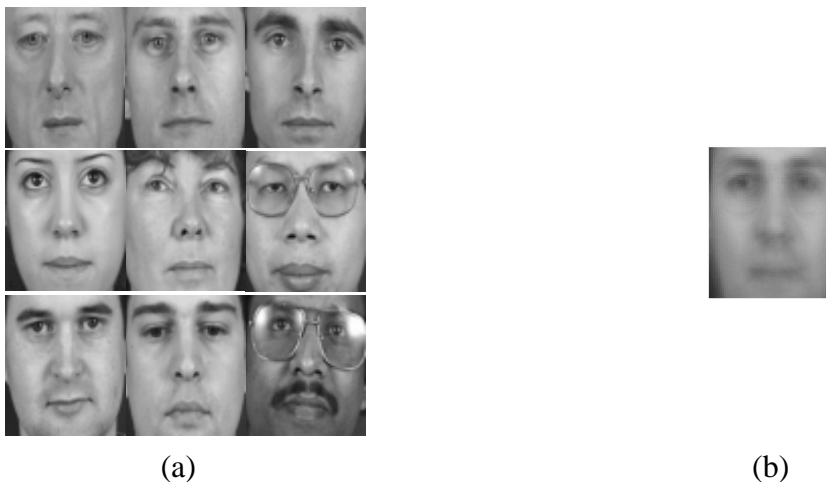


Figure 2.4 : a) Exemple d'images utilisées dans l'ensemble d'apprentissage, b) Visage moyen.

La matrice de covariance des données C est calculée suivant :

$$C = AA^T = \frac{1}{N} \sum_{i=1}^N A_i A_i^T \quad (2.10)$$

où $A = [A_1, A_2, \dots, A_N]$.

Le nouvel espace réduit de visage, qui constitue une base orthogonale, est déterminé en résolvant le problème suivant :

$$C \Phi = \Phi \Lambda \quad (2.11)$$

$\Phi = [\phi_1, \dots, \phi_n]^T$ est la matrice des vecteurs propres de C , et Λ est la matrice diagonale avec les valeurs propres $\lambda_1 \geq \dots \geq \lambda_n$ de C sur sa diagonale.

La matrice C est généralement de grande taille ($n*m, n*m$) et le calcul des vecteurs propres devient très vite fastidieux. Turk et Pentland [TP91] considèrent la matrice $J = A^T A$ de dimension (N, N) . Soit v_i un vecteur propre de J .

$$J v_i = \lambda_i v_i \quad (2.12)$$

C'est à dire

$$A^T A v_i = \lambda_i v_i \quad (2.13)$$

En multipliant par A

$$A A^T A v_i = \lambda_i A v_i \quad (2.14)$$

C'est à dire

$$C A v_i = \lambda_i A v_i \quad (2.15)$$

Ainsi pour retrouver les vecteurs propres de C il suffit de calculer les vecteurs propres de J et de les multiplier par la matrice des données A . Quelques exemples de vecteurs propres, appelés *Eigenfaces*, redimensionnés en images de taille (n, m) sont illustrés sur la figure 2.5.



Figure 2.5 : Les six premiers *Eigenfaces* calculés à partir de l'ensemble d'images de la figure 2.4.

Le premier système réussi d'identification de visages a été construit par Turk et Pentland [TP91] en utilisant les *Eigenfaces* pour la détection et l'identification de visage. Etant donné les K premiers vecteurs propres, correspondant aux plus grandes valeurs propres, chaque visage dans la base de données est représenté comme un vecteur de poids $\Omega_i = [w_1, w_2, \dots, w_K]$ obtenu en projetant l'image sur la matrice des vecteurs propres par un simple produit. Quand une nouvelle image de test dans laquelle on cherche à identifier un visage est donnée, celle-ci est également représentée par son vecteur de poids Ω . L'identification de l'image de test est réalisée en localisant l'image dans la base de données dont les poids sont les plus proches des poids de l'image en question au sens d'une mesure de similarité.

En remarquant que les projections d'une image de visage et d'une image ne contenant pas de visage sont remarquablement différentes, une méthode pour détecter la présence d'un visage dans une image est déduite [TP91]. Etant donné une image inconnue Γ_l , les auteurs calculent la distance d (équation 5) qui sépare l'image centrée avec sa projection sur l'espace de visage construit. Si d est inférieure à un certain seuil T_d alors Γ_l est considéré comme un visage.

$$d = \left\| A_l - \sum_{i=1}^K (\Phi_i \Phi_i^T A_l) \right\| \quad (2.16)$$

Nous avons traité dans notre cas des images brutes de visages de la base XM2VTS mais dont la position précise de ceux-ci est inconnue. Pour localiser un visage dans une image donnée, nous avons fait parcourir une fenêtre sur toute l'image avec un pas de 5 pixels et nous avons mesuré pour chaque position la distance d relative. La fenêtre qui induit la distance minimale est ainsi inférée comme la position exacte du visage. La figure 2.6 illustre quelques exemples de localisations.



Figure 2.6 : Exemple de localisation par la méthode des *Eigenfaces*.

Nous remarquons principalement la limite de cette méthode à localiser des visages de plus petites ou de plus grandes tailles que les images qui ont servi à l'apprentissage. D'autre part, le rectangle localisé n'épouse pas la forme exacte du visage traité. Nous proposons dans la section suivante une méthode originale de localisation adaptée à chaque visage traité par les réseaux de neurones.

2.5.3 Méthode de localisation par réseau de neurones

Le réseau de neurones est utilisé ici pour apprendre les informations contenues dans les moments de Zernike, respectivement les coefficients *Eigenfaces*, et leurs relations avec le contour du visage. Il est à noter ici, qu'on ne fait aucune supposition sur la forme probable du visage recherché de même qu'on ne fait subir à l'image aucun prétraitement.

2.5.3.1 Organisation du réseau

Un réseau de neurones est composé d'un ensemble de neurones formels² interconnectés entre eux donnant naissance à des réseaux à structures variées. Pour notre application, nous avons utilisé la structure en couches successives (*Multi-layer Perceptron : MLP*). Une telle structure (cf. figure 2.7) diffuse l'information de la couche d'entrée, composée par les neurones formels recevant les informations primitives, vers la couche de sortie, qui contient les neurones finaux transmettant les informations de sortie traitées par la totalité du réseau, tout en traversant une ou plusieurs couches intermédiaires, dites couches cachées [Vap00]. Le réseau ainsi établi est un système non linéaire qui associe, aux vecteurs caractéristiques d'entrée, le contour du visage de la couche de sortie.

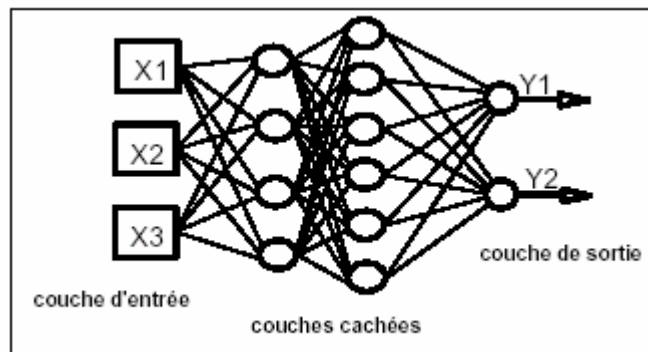


Figure 2.7. Réseau de neurones multicouche (Perceptron Multicouche PMC ; en anglais Multi-layer Perceptron MLP).

Le réseau de neurones développé utilise une fonction de seuillage sigmoïdale et une rétropropagation du gradient pour optimiser sa phase d'apprentissage. La rétropropagation consiste à rétropropager l'erreur commise par un neurone à ses synapses et aux neurones qui y sont reliés. Pour les réseaux de neurones, on utilise habituellement la rétropropagation du gradient de l'erreur, qui consiste à corriger les erreurs selon l'importance des éléments qui ont justement participé à la réalisation de ces erreurs : les poids synaptiques qui contribuent à engendrer une erreur importante se verront modifiés de manière plus significative que les poids qui ont engendré une erreur marginale.

2.5.3.2 Mise en œuvre de la localisation

La mise en œuvre de la méthode repose essentiellement sur la phase d'apprentissage du réseau (voir figure 2.8) et se résume en quatre étapes :

- Calcul des vecteurs moments de Zernike (respectivement les coefficients *Eigenfaces*) pour l'ensemble des N images de la base de travail.
- Construction de la base d'apprentissage par prélèvement aléatoire de M images de la base ($M \ll N$) et des M vecteurs Zernike Z_i (respectivement vecteurs *Eigenfaces* E_i) leurs correspondant.

² Un neurone formel est une représentation mathématique et informatique d'un neurone biologique. Le neurone formel possède généralement plusieurs entrées et une sortie qui correspondent respectivement aux dendrites et au cône d'émergence du neurone biologique (point de départ de l'axone). Les actions excitatrices et inhibitrices des synapses sont représentées, la plupart du temps, par des coefficients numériques (les poids synaptiques) associés aux entrées. Les valeurs numériques de ces coefficients sont ajustées dans une phase d'apprentissage.

- Délimitation manuelle de la zone du visage dans chacune des images de la base d'apprentissage par un ensemble de points représentant le contour C_i de forme adaptée à chaque visage traité.
- Entraînement du réseau sur les exemples de couples (Z_i, C_i) , respectivement (E_i, C_i) .

Les figures 2.8 et 2.9 explicitent les deux phases d'entraînement et de test pour la localisation du contour du visage.

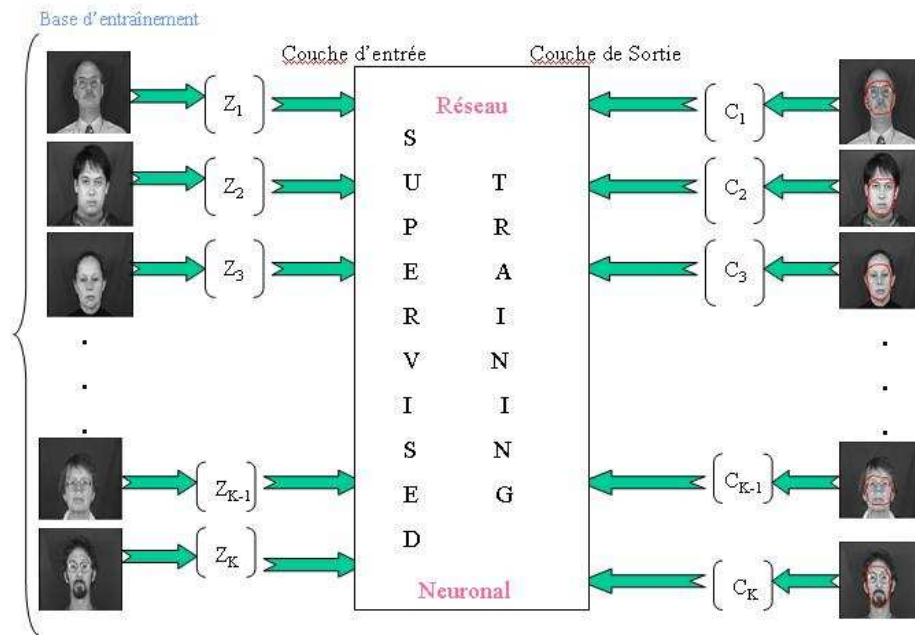


Figure 2.8 : Phase d'apprentissage de la procédure de localisation faciale par réseau de neurones.

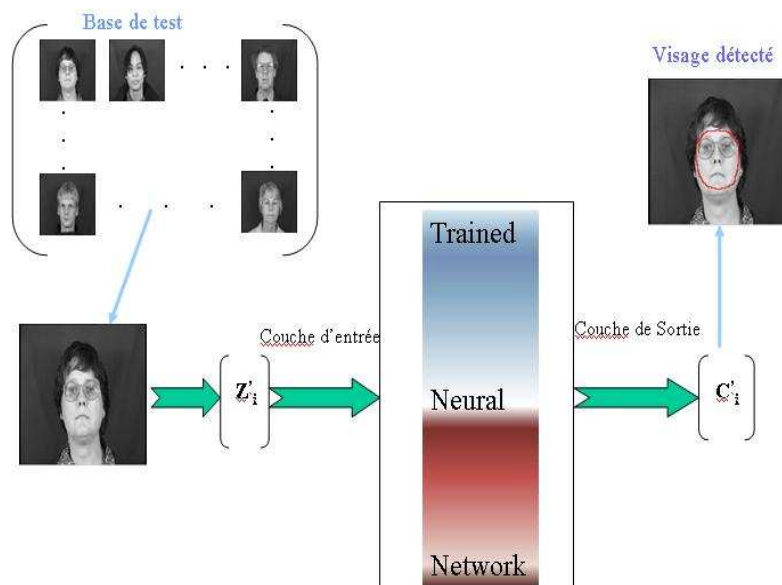


Figure 2.9 : Phase de test de la procédure de localisation faciale par réseau de neurones.

Pour avoir une bonne aptitude de généralisation, les exemples d'entraînement doivent être assez représentatifs de la base de travail. Pour tester les performances du réseau obtenu on procède alors sur l'ensemble des $(N-M)$ images restantes dans la base de travail.

2.5.4 Qualité de la Localisation

Pour le calcul des performances, nous avons développé un critère de mesure qui permet de donner une évaluation objective de la qualité de la localisation obtenue à la sortie du système.

2.5.4.1 Etat de l'art

La localisation faciale, consiste à trouver la position exacte du visage dans une image. Malheureusement, il est difficile de mesurer les performances d'un algorithme de localisation par défaut de critère de qualité universel qui formalise cette tâche [RCBM06]. Ainsi, la comparaison directe des systèmes de localisation reste très subjective du fait qu'ils fournissent des « taux » de localisation et d'erreurs sans expliciter comment une tâche de localisation est appréciée comme correcte ou incorrecte. Souvent une fenêtre détectée par le système est comptée comme valide par observation visuelle si la fenêtre inclut les yeux, le nez et la bouche [Hua98]. Peu de travaux donnent une précision sur ce critère. Dans [LKP03] par exemple, Lienhart et al. décident qu'une localisation est correcte si la distance euclidienne entre les centres du visage détecté et du vrai visage est inférieure à 30% de la largeur du vrai visage, et la largeur du visage détecté est entre $\pm 50\%$ de celle du vrai visage. Selon Garcia et Delakis [GD04] : un visage est considéré comme correctement détecté si la taille de la fenêtre englobante n'excède pas plus de 20% de la taille réelle de la région faciale, et qu'elle contient les yeux, le nez et la bouche. Dans [FK02] les auteurs considèrent une bonne localisation si les mesures de la position (à travers la position des yeux) et la taille (à travers la distance entre les yeux) du visage ne dépassent pas 30% des vraies valeurs. Jesorsky et al. [JKF01] ont introduit une mesure d'erreur basée sur la distance de Hausdorff entre les vrais centres des yeux et ceux détectés.

2.5.4.2 Proposition d'un critère de localisation

Nous proposons ici de mettre en œuvre une mesure objective basée sur le rapport entre la surface de visage détecté et la surface de visage présent dans l'image originale. Pour permettre de réaliser une telle mesure, l'ensemble des images de la base de test ont été divisées en trois régions (cf. figure 2.10). La région blanche contient les W pixels jugés essentiels dans la détection du visage (nez, yeux, bouche,...). La région grise contient des pixels éléments du visage mais jugés non essentiels pour sa bonne détection. La région noire contient les B pixels non éléments du visage.

La région blanche doit être impérativement contenue dans le contour résultant d'un algorithme de détection. La détection de pixels de la région grise est considérée comme optionnelle (subjective) et n'influera pas sur le résultat de la mesure. Par contre aucun pixel contenu dans la région noire ne devrait être détecté, par conséquent, si un pixel élément de cette région est détecté, cela pénalisera le résultat de la mesure.



Figure 2.10 : Images originales et les trois zones caractéristiques associées à chacune d'elles.

Le critère de mesure proposé est basé sur le calcul des deux quantités nommées Gdr (*Good detection rate*) et Qdr (*Quality detection rate*) définies par les équations (2.17) et (2.18) :

$$Gdr = \frac{W_1}{W} \cdot 100 \quad (2.17)$$

$$Qdr = \left(\frac{W_1}{W} - \frac{B_1}{A - B} \right) \cdot 100 \quad (2.18)$$

W_1 et B_1 représentent respectivement le nombre de pixels correctement et faussement détectés comme faisant partie du visage et A le nombre total des pixels de l'image d'origine. Gdr mesure à quel point les pixels composant l'essentiel du visage ont été correctement détectés. Qdr donne une mesure plus stricte prenant en compte le nombre de pixels faussement détectés comme appartenant au visage dans le calcul de la valeur quantitative de détection. Ces deux mesures sont complémentaires. En effet, si on connaît seulement Gdr on ne possède aucune information sur le taux de pixels attribués à tort au visage. Par ailleurs, la seule connaissance de Qdr ne donne pas d'information sur le nombre de pixels détectés appartenant à l'essentiel du visage et qui ne sont pas inclus dans la région détectée comme visage. En conclusion un algorithme de détection sera d'autant meilleur que Gdr sera élevé et Qdr proche de Gdr. Sur la figure 2.11 nous donnons quelques exemples de détection qui montrent la complémentarité de ces deux mesures Gdr et Qdr.

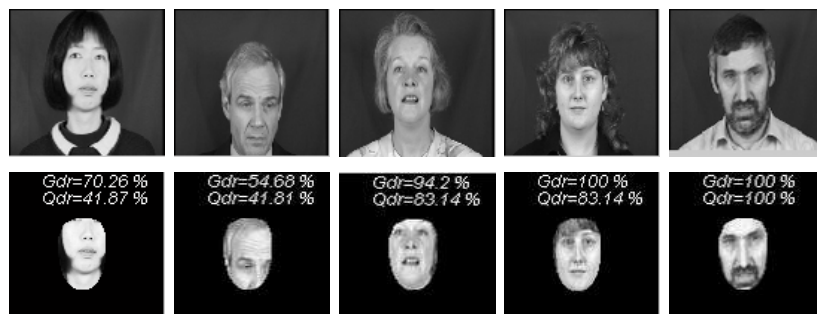


Figure 2.11: Exemples sur la complémentarité des deux mesures Gdr et Qdr.

2.6 Résultats expérimentaux

Afin de valider notre méthode nous avons utilisé la base XM2VTS [Mes99] qui contient 295 individus avec 4 sessions et 2 images par session pour chaque individu soit un total de 8 images par personne. Ces images contiennent des variations de luminosité ainsi que de petites variations de poses avec un hochement horizontal et vertical de la tête comme le montre la figure 2.11.

Dans notre protocole d'expérimentation nous avons pris pour chaque personne quatre images de différentes sessions ce qui fait une base de travail de 1180 images. Nous avons ensuite transformé les images en niveau de gris et nous les avons réduites à 200x200 pixels pour le calcul des *Eigenfaces* et des moments de Zernike.

Pour construire la base d'apprentissage, nous avons effectué un choix aléatoire de 15 personnes différentes (à partir des 295 que contient la base) pour sélectionner 3 poses de chaque personne, ce qui nous donne 45 exemples ($\mathbf{Z}_i, \mathbf{C}_i$) ou ($\mathbf{E}_i, \mathbf{C}_i$) pour l'entraînement. Pour avoir une idée précise et assez générale sur les performances de la méthode, nous avons procédé à la construction de 20 bases d'apprentissage par prélèvement aléatoire des exemples. L'ensemble de test est le même pour toutes ces configurations d'apprentissage. En effet, nous avons pris l'image de la quatrième session de chaque individu dans la base de test. Notre travail d'expérimentation a visé l'étude du comportement de la méthode vis à vis de la base d'apprentissage, de la taille des vecteurs caractéristiques et de la complexité du réseau de neurones utilisé. L'estimation de la qualité de la localisation se fait à travers la procédure quantitative décrite précédemment.

Le réseau de neurones utilisé lors de nos expériences est organisé suivant une structure multicouche (MLP) avec une seule couche cachée. Pour cette dernière couche, nous avons retenu un nombre de 10 neurones permettant de donner des résultats légèrement meilleurs qu'avec des couches formées par 5, 15, 20 ou 25 neurones formels. D'autre part, la couche d'entrée est formée par un nombre de neurones qui correspond à la taille des vecteurs caractéristiques (entrés du réseau). La couche de sortie, quant à elle, est formée par 60 neurones correspondant aux coordonnées x et y des 30 points du contour localisant le visage.

2.6.1 Etude comparative de la caractérisation Moments de Zernike - *Eigenfaces*

Tout d'abord nous présentons sur la figure 2.12 le résultat de l'application de la méthode proposée à une base de test formée par les 295 images représentant le quatrième enregistrement de chaque individu de la base XM2VTS et qui n'a pas été introduit lors de la phase d'apprentissage.

Ces résultats ont été obtenus à travers l'apprentissage d'un réseau à 22 neurones sur la couche d'entrée. L'allure générale des deux courbes montre que la plupart des visages ont été assez bien détectés (Gdr et Qdr supérieurs à 70%).

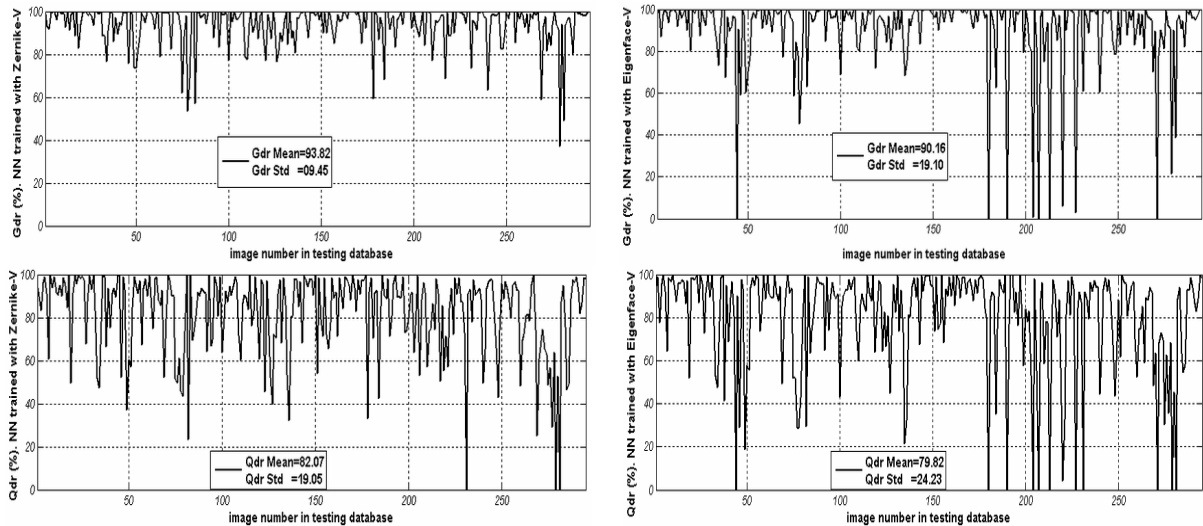


Figure 2.12 : Courbes des taux de localisation Gdr (en haut) et Qdr (en bas) pour chaque image test. Ces taux sont obtenus par un réseau de neurones entraîné par les moments de Zernike (à gauche) et les *Eigenfaces* (à droite).

Nous présentons sur la figure 2.13 un exemple de résultats obtenus par application de deux réseaux de neurones, séparément entraînés l'un sur des vecteurs de moments de Zernike et l'autre sur des vecteurs *Eigenfaces*, sur le même ensemble d'images de la base XM2VTS.



Figure 2.13 : Résultats de localisation et valeurs de Gdr et Qdr obtenus pour des images de la base XM2VTS. En haut, images d'origine. Au milieu, localisation par l'algorithme utilisant les moments de Zernike. En bas, localisation par les *Eigenfaces*.

Ces résultats ont été obtenus pour des vecteurs caractéristiques (*features*) de dimension 22. Les taux de localisation enregistrés sur les deux types de courbes montrent la supériorité de performances en faveur de la caractérisation par moments de Zernike. Le tableau 2.1 illustre le bon comportement des moments de Zernike face aux *Eigenfaces*. En effet, on constate que Gdr et Qdr sont plus importants en moyenne lorsque l'image est caractérisée par les moments de Zernike. De même les écart-types plus faibles obtenus permettent d'illustrer le meilleur centrage des résultats autour de ces valeurs moyennes.

	moments de Zernike		<i>Eigenfaces</i>	
	Nombre d'images	%	Nombre d'images	%
Gdr<70%	11	3.72%	23	7.79%
Qdr<70%	60	20.33%	74	25.08%
Gdr <i>Mean / Std</i>	93.82% / 9.45%		90.16% / 19.10%	
Qdr <i>Mean / Std</i>	82.07% / 19.05%		79.82% / 24.23%	

Tableau 2.1 : Résultats des deux mesures de qualité de localisation de visage obtenus par les moments de Zernike et les *Eigenfaces*

Sur le tableau 2.1 nous montrons que pour 96,28% des images caractérisées par les moments de Zernike l'essentiel du visage (Gdr>70%) est dans la région localisée. Ce taux de détection est de 92,21% pour les *Eigenfaces*. Cependant, pour donner des résultats probants en terme de détection, nous adoptons un critère assez rigoureux pour estimer une bonne localisation de visage, à savoir un Gdr et un Qdr supérieurs à 70%. Si on retient ce dernier seuil comme critère de bonne localisation, notre taux de détection du visage sera de 80% par moments de Zernike contre 75% pour les *Eigenfaces*. L'écart-type des valeurs moyennes enregistrées montre la capacité de généralisation du réseau.

Les moments de Zernike possèdent, ainsi, une meilleure capacité de généralisation avec 9,45% d'écart-type sur la moyenne Gdr contre 19,10% pour les *Eigenfaces*. D'autres expérimentations portant sur l'influence de la base et de la taille des vecteurs d'entraînement ont été réalisées [SCL⁺07]. La comparaison de performances selon l'influence de la base d'entraînement a été étudiée sur 20 bases aléatoirement construites. Les résultats obtenus confirment la supériorité des performances des moments de Zernike et ce pour la plus part des bases d'entraînement utilisées. D'autre part, en employant des vecteurs caractéristiques de dimension différente, la même constatation a été faite sur la différence des taux de détections enregistrées en faveur des réseaux entraînés par les moments de Zernike.

2.7 Normalisation des visages

Nous décrivons dans cette section le processus de normalisation que nous avons appliqué sur les images de visages en amont de la reconnaissance. La normalisation consiste à aligner tous les visages de la même manière dans leurs images associées. Elle est nécessaire pour extraire efficacement les similarités et les dissimilarités des visages photographiés et garantir de bonnes performances de la reconnaissance faciale. Nous développons, dans ce paragraphe, la procédure classique de normalisation basée sur les centres des yeux. Pour mener à bien cette étape de normalisation, nous disposons pour la plupart des bases d'images que nous avons utilisées des positions des yeux. Pour le reste des images, nous avons annoté manuellement ces points d'intérêt. Quelques récents travaux proposent d'autres méthodes de normalisation automatiques sans initialisation par les centres des yeux. Les auteurs dans [WAH03] utilisent les modèles d'apparence *AAM* pour normaliser en deux temps le visage détecté. La première étape utilise un modèle simple d'apparence (Simple Direct Appearance Model - *SDAM*) pour localiser quelques points du visage en s'appuyant sur la texture. La deuxième étape élabore une localisation fine en utilisant une variante de la méthode *AAM*.

Généralement la normalisation consiste en une rotation du visage dans l'image de manière à ce que l'axe interoculaire soit horizontal, suivi d'un centrage du visage dans l'image. Tous les visages doivent être représentés à la même échelle, et les images normalisées sont toutes de même taille. Pour cela, nous avons redimensionné les visages de notre base de travail à l'aide d'un filtrage bilinéaire pour avoir une distance interoculaire égale à 36 pixels. Cette dernière valeur est choisie pour aboutir à une image normalisée de taille 64 x 64 pixels tout en assurant des proportions adéquates à gauche et à droite des centres des yeux. A la suite de cette mise à l'échelle de l'image, le découpage est réalisé de manière à ne retenir que la région faciale. Les proportions de un quart et de trois quart de la taille de l'image sont retenues respectivement au dessus et au dessous de l'axe interoculaire (voir figure 2.14).

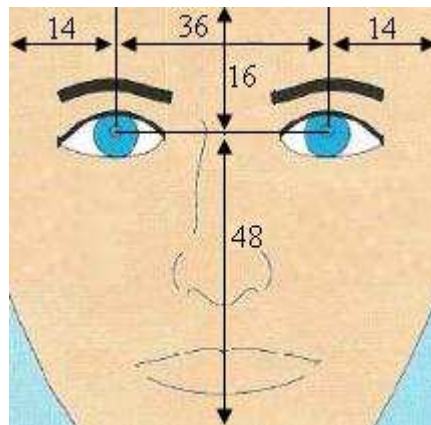


Figure 2.14 : Nombres de pixels retenus pour le découpage des images de visages

Les images de couleur sont transformées en images à 256 niveaux de gris en moyennant les valeurs des trois canaux RGB. En cas de besoin (images sombres dues à un manque de luminosité ou bien trop éclairées), 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. Ainsi cette opération vise à harmoniser les valeurs de pixels des différentes images, afin de gommer les dissimilarités dues à des différences de conditions d'éclairage par exemple. Nous illustrons ci-dessous les différentes étapes du processus de normalisation utilisé dans le cadre de cette thèse.



Figure 2.15 : Le processus de normalisation des images compte quatre étapes : la rotation, le redimensionnement, le découpage, et l'égalisation d'histogramme.

2.8 Conclusion

Ce chapitre nous a permis de détailler la première étape d'un processus de reconnaissance automatique de visages, à savoir la détection/localisation/normalisation de la face. Nous avons pu voir, à travers la présentation des deux méthodes développées, à savoir les moments de Zernike et les *Eigenfaces* [SCL⁺07], que la détection parfaite du visage était un objectif difficile à atteindre.

La comparaison des résultats obtenus au cours de ces deux expérimentations a été effectuée sur la base des critères de qualité élaborés par M. Saaidia. La localisation de 30 points d'un contour de visage par un réseau de neurones entraîné par l'une ou l'autre des méthodes est à l'avantage des moments de Zernike pour élaborer le vecteur d'entrée du réseau. Le visage une fois détecté, une procédure classique de normalisation est appliquée. Meilleure est la précision de détection du visage, plus élevées sont les performances des systèmes de reconnaissance dont nous allons faire un tour d'horizon dans le chapitre suivant.

Chapitre III

Reconnaissance faciale - Etat de l'art des techniques de reconnaissance

3.1 Introduction

La reconnaissance automatique du visage est la technique la plus commune et populaire parmi les techniques de reconnaissance biométrique. Elle reste la plus acceptable puisqu'elle correspond à ce que les humains utilisent dans l'interaction visuelle. Elle reçoit une attention accrue du fait de son caractère non invasif, au sens où elle ne requiert pas la coopération du sujet. Le développement d'outils de suivi (tracking) performant a permis son déploiement à large échelle. Ainsi, la reconnaissance du visage est utilisée dans un grand nombre d'applications incluant la sécurité, la vérification de personnes pour le contrôle d'accès, la communication par Internet et le loisir par ordinateur. Bien que les recherches en reconnaissance automatique de visage aient été conduites depuis plus de trois décennies, ce problème reste encore non résolu. Les dernières années ont vu des progressions significatives dans ce domaine du fait des avancées dans la modélisation des visages, des techniques d'analyses et des technologies d'évaluation. Des systèmes ont été développés pour la détection et le "tracking" des visages mais la reconnaissance efficace de visage offre encore un grand challenge pour les chercheurs en vision par ordinateur et reconnaissance de forme.

Ce chapitre est organisé comme suit. Après avoir positionné notre champ de traitement sur la reconnaissance faciale 2D parmi tant d'autres domaines de reconnaissance de visage dans la section suivante, nous présentons dans la section suivante le processus de reconnaissance humaine de visage, afin de tirer les meilleurs enseignements des excellentes aptitudes du cerveau humain dans ce domaine. La troisième section met en lumière les difficultés majeures liées à la reconnaissance automatique de visage. Enfin dans la dernière section, nous détaillons l'état de l'art des méthodes de reconnaissance de visage 2D.

3.2 Positionnement

La plupart des méthodes de reconnaissance automatique de visage portent sur la représentation des images fixes 2D. On dispose d'une base de données, appelée communément galerie, contenant des photographies de personnes connues (dont l'identité est connue) et 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 galerie. Le système de reconnaissance est généralement basé sur un modèle construit à partir d'une base d'apprentissage contenant un ensemble d'images de visages représentatifs et généralisable aux images requêtes. Dans certains cas, la base d'apprentissage et la galerie sont confondues. Le modèle est alors spécifiquement conçu pour les visages qu'il vise à reconnaître. Dans

d'autre cas au contraire, soit du fait des spécificités de la technique utilisée, soit à cause de l'application, la base d'apprentissage et la galerie sont distinctes.

La plupart des systèmes sont basés sur l'utilisation des images à niveau 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 [SL05, Gut01].

Depuis la fin des années 1990, suite à l'amélioration des caméras numériques permettant d'avoir des vidéos de meilleures qualités, quelques équipes de recherche se sont focalisées sur la reconnaissance de visages à partir de séquences vidéo [LGL01, Zha03, PVY03, CCZ05]. Les premiers systèmes reposent sur l'utilisation des techniques initialement conçues pour les images fixes. Le système détecte automatiquement une image représentative de l'ensemble des vues du visage requête et se ramène à un problème de reconnaissance d'images fixes. D'autres techniques utilisent un algorithme de vote majoritaire sur les résultats de classification de l'ensemble des vues du visage requête [GMP00]. Une amélioration de ce type de système s'est développée grâce à la mise au point des techniques de suivi des visages. Une vue frontale virtuelle est synthétisée via l'estimation de la pose et de la profondeur à partir vidéo. Plus récemment, d'autres méthodes exploitent d'une manière cohérente l'information spatiale contenue dans chaque vue et l'information temporelle telle que la trajectoire des points caractéristiques du visage par exemple, pour construire un modèle spécifique de chaque personne de la galerie [LC03, LK05, AC06]. 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.

Très récemment, les avancées dans le domaine de l'acquisition de données tridimensionnelles, notamment par le biais de scanner 3D, ont permis l'émergence de technologies de reconnaissance 3D basées principalement sur la forme. La motivation principale de ces techniques est de pallier la sensibilité inhérente des techniques de reconnaissance 2D quant aux variations d'éclairage et de pose. Cependant, ces méthodes de reconnaissance 3D restent très sensibles à la variabilité intrinsèque du visage produite par les changements d'expressions faciales et l'évolution de la physiologie, ce qui constitue le plus grand challenge pour cette technologie. Il existe deux grandes familles de méthodes. Les méthodes basées sur la mise en correspondance directe d'objets 3D représentés par des points tridimensionnels, des courbes de niveau, des courbes géodésiques ou des fonctionnelles de courbures [Amo06]. La deuxième famille repose sur les méthodes de projection sur des sous espaces (*subspace methods*) développé initialement pour l'analyse et la reconnaissance des visages 2D [PW05].

Dans le cadre de cette thèse, nous nous intéressons aux techniques basées sur l'étude d'images 2D à niveau de gris de visages. La plupart des méthodes étudiées 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 être étendues à la reconnaissance dans les séquences vidéo.

3.3 La reconnaissance humaine de visages

Les travaux portant sur la reconnaissance humaine des visages ont été intensivement menés dans les domaines de la psychologie et de la neurophysiologie durant ces trois dernières décennies. C'est avec un grand intérêt que les scientifiques cherchent à élucider les prodigieuses facultés de reconnaissance du cerveau humain afin de concevoir un algorithme de reconnaissance automatique. Nous présentons dans ce qui suit un bref aperçu de ces

recherches afin de tirer les meilleurs enseignements des excellentes aptitudes du cerveau humain dans ce domaine. Un inventaire assez complet des principales études relatives à la reconnaissance humaine des visages est dressé dans [Zha03, O'T04, Wec06].

Le visage humain est une collection de deux os (le crâne et la mâchoire) et de 44 muscles. Ceci donne une grande liberté de mouvement qui permet des milliers de différentes expressions faciales incluant les mouvements effectués lors de la parole, le changement du point de vue ou de la pose de la tête, et les expressions comme le sourire, le dégoût, la peur ... Ces mouvements du visage jouent un rôle important dans l'interaction sociale des individus. En effet, reconnaître les visages est indispensable à notre humanité. Il est estimé que 50% de notre fonction cervicale est consacrée à la vision et que la grande part de celle-ci va pour la reconnaissance faciale [Wec06]. Ainsi, l'aptitude à reconnaître ses semblables est l'une des capacités cognitives les plus importantes de la race humaine. La reconnaissance d'une personne par le cerveau humain passe par l'utilisation de nombreuses informations visuelles (visage, posture, lieu de rencontre, etc.) ou non visuelles (auditives (voix), olfactives (parfum), etc.) obtenu par tous les organes sensoriels. Dans plusieurs situations, le contexte est aussi un facteur important : on reconnaîtra plus rapidement un étudiant dans sa classe 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 en une fraction de seconde.

3.3.1 La reconnaissance humaine de visages : un processus complexe

Les humains sont dotés d'une excellente aptitude à identifier leurs semblables. Les études neurobiologiques tendent à prouver que la reconnaissance humaine des visages constitue un processus qui est mené dans une région particulière du cerveau. Gauthier et Logothetis [GL00] ont montré que le processus engagé par le cerveau pour la reconnaissance faciale n'est pas différent des autres tâches de reconnaissance d'objets. Cependant une sélection de cellules spécifiques est impliquée pour la reconnaissance de chaque classe d'objet. Ainsi, ils ont prouvé qu'une sélection de cellules néocorticales et limbiques dans le cortex préfrontal est activée en réponse à l'identité faciale, qu'une autre sélection de cellules dans la scissure supérieure est activée pour la reconnaissance de la pose, et que des cellules dans l'amygdale sont activées pour reconnaître l'expression faciale. Quant à l'analyse des caractéristiques globales du visage, une sélection de toutes les cellules mentionnées ci-dessus est activée.

La reconnaissance automatique des visages constitue un domaine particulier de la reconnaissance de forme générique. En effet, les visages de deux personnes différentes sont structurellement très proches, car dotés des mêmes caractéristiques faciales dont la localisation varie très peu. De plus, les sources de variabilités 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, cette tâche qui permet de classer un visage dans la catégorie des visages est appelée détection de visage. En revanche, dans le cadre de la reconnaissance, nous connaissons la nature de l'objet, mais nous 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.

3.3.2 Processus local et global

Chaque visage humain est unique et fournit des informations sur son identité. Les humains peuvent mémoriser des centaines (si ce n'est des milliers) de visages différents. Ceci dépasse de loin notre capacité de mémoriser différents exemples de n'importe quelle autre classe d'objets. L'information qui rend les visages uniques doit être trouvée dans de subtiles variations de la forme et de la configuration globale des caractéristiques faciales (yeux, nez, bouche). Des expérimentations neuropsychologique [Zha03] suggèrent que les individus utilisent les caractéristiques globales et locales pour reconnaître les visages de leurs semblables avec une éventuelle tendance de dépendance à la configuration globale des caractéristiques. L'analyse de l'information des basses fréquences de l'image, reflétant sa globalité, est suffisante à la reconnaissance d'un visage familial. En effet, les bébés ont une vision floue à la naissance, mais les nouveaux nés peuvent distinguer, en utilisant les caractéristiques globales, le visage de leur mère parmi les autres visages peu après la naissance. Par contre, les cheveux, la forme du visage, les yeux et la bouche sont reconnues comme étant des caractéristiques primordiales dans la perception et la mémorisation de nouveaux visages. D'autres études montrent que l'importance du nez est mineure dans le cadre de la reconnaissance d'un visage de face, mais devient importante pour une pose de profil, où l'on peut mieux appréhender sa forme. Il a également été montré que la partie supérieure du visage a une plus grande importance que la partie inférieure [BHB98]. Ces études plaident en faveur des techniques hybrides pour la reconnaissance automatique des visages, au sens où elles sont basées sur l'étude conjointe des caractéristiques globales et locales des visages.

3.3.3 Impact des différents types de variabilités

Il est également très intéressant d'étudier l'impact des différentes sources de variabilité sur les performances du système visuel humain en terme de reconnaissance de visage. 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 [Vis05].

La direction d'éclairage influe sur l'aptitude humaine à reconnaître un visage. Les conditions d'éclairage les plus favorables consistent en un éclairage par le haut [HB96]. Johnston et al. [JHC92] ont montré qu'un visage familial éclairé par une source lumineuse situé sous le visage est plus difficile à identifier. Par contre, l'expression faciale a peu d'influence sur nos capacités de reconnaissance, pour autant qu'elle reste raisonnable. D'autre part, 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 quart cette tâche est relativement aisée. 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].

Ces études illustrent l'importance de l'apprentissage pour la reconnaissance humaine. Nous avons, depuis notre enfance, observé suffisamment de visages pour avoir complètement intégré leurs symétries et être capable d'inférer une vue à partir de son opposée. A 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 que les dissimilarités entre deux visages différents ainsi que les variations possibles entre deux vues d'un même visage sont nécessaire pour construire un outil de reconnaissance automatique performant.

3.4 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 influencer 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 [DGB08].

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, 2002 et 2006 [BBP01, PGM⁺03, Phi07] et par Gross et al. [GBMK04, GSC01] nous permettra de tirer un certain nombre de conclusions. Le protocole FRVT a permis de caractériser essentiellement l'impact de la pose, des conditions d'éclairage, du délai entre différentes prises de vue, de facteurs individuels et de la taille de la base comme nous l'avons exposé en section 1.6. Dans [GSC01], Gross et al. fournissent une étude systématique de l'impact de différents paramètres sur les performances du système. Les six facteurs considérés sont : la pose de la tête, les changements d'éclairage, l'expression faciale, les occultations, l'intervalle de temps entre deux prises de vue et le sexe (genre). Leurs influences relatives sont étudiées en les faisant varier isolément ou conjointement.

3.4.1 Les variations de la pose

Un changement de l'angle d'inclinaison du visage engendre, pour une position fixe du capteur, de nombreux changements d'apparence dans l'image collectée. Nous nous intéressons ici aux rotations du visage en profondeur tels que les mouvements de type hochement de tête ou négation. En effet, la phase préliminaire de normalisation du visage détaillée en section 2.7 permet de corriger d'éventuelles rotations dans le plan de l'image. Les rotations en profondeur engendrent l'occultation de certaines parties du visage comme pour les vues de trois quarts. D'autre part, elles amènent des différences de profondeur qui, projetées sur le plan 2D de l'image, résultent en des déformations qui font varier la forme globale du visage. Ces déformations qui consistent à l'étirement de certaines parties du visage et le compactage d'autres régions font varier aussi les distances entre les caractéristiques faciales.

Si la pose du visage-requête diffère significativement de celle des visages enregistrés, les performances des systèmes de reconnaissance baissent sensiblement. En effet, selon le FRVT 2000 [BBP01], la rotation de la tête n'entraîne pas de baisse significatives des taux de reconnaissance jusqu'à $\pm 25^\circ$. Ce seuil, dépassé, on constate une chute de 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 actuels sont de l'ordre de 90%. Des rotations plus importantes engendrent une forte baisse des performances.

3.4.2 Les changements d'éclairage

L'intensité et la direction d'éclairage lors de la prise de vue influent énormément sur l'apparence du visage dans l'image. Dans la plupart des applications réelles, des changements dans les conditions d'éclairage sont néanmoins inévitables, notamment lorsque les vues sont collectées à des dates différentes, en intérieur ou en extérieur. Etant donné la forme

intrinsèque d'un visage humain, des changements d'éclairage peuvent faire apparaître sur le visage des ombres accentuant ou masquant certaines caractéristiques faciales. L'évaluation FRVT [BBP01] conclue que des changements importants dans les conditions d'éclairage peuvent mener à des baisses considérables dans les taux de reconnaissance. En effet, si la plupart des systèmes de reconnaissance de visages sont stables à des changements raisonnables des conditions d'éclairage en intérieur, on note en extérieur des déficits de la reconnaissance de l'ordre de 40%. Ces baisses de taux de reconnaissance peuvent être dues à des variations de la luminosité réfléchiée par la peau mais aussi 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. Les auteurs ont montré que les algorithmes actuels de reconnaissance de visages sont assez robustes aux effets de changements d'éclairage purs, en partie corrigés par une égalisation d'histogramme appliquée lors de la phase de normalisation. Ce serait donc les ajustements non linéaires des caméras ainsi que l'éclairage non homogène du visage qui engendreraient en grande partie la réduction de performance enregistrées ce qui nécessite des corrections plus sophistiquées.

3.4.3 Les expressions faciales

Les visages sont des objets non rigides. L'expression faciale de l'émotion, combinée avec la parole, peut produire des changements d'apparence importants des visages. Le nombre de configurations possibles est incalculable. 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] soutiennent que les expressions faciales n'ont pas une grande influence sur les algorithmes de reconnaissance, pour autant qu'elles restent raisonnables. Alors que les cas extrêmes qui engendrent d'importantes déformations de la bouche (tel que le cri) et le rétrécissement ou la fermeture complète des yeux dégradent considérablement les performances de la reconnaissance automatique.

Il peut 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 [Abb04]. Soit plusieurs modèles de visages ont été appris, un par catégorie d'expression faciale [FL03, BLFM03]. Dans ce cas, il serait possible de comparer directement le visage-test à la base des visages arborant la même expression. Par contre, à l'aide d'une technique générative utilisant un modèle de visage suffisamment précis, il serait possible de transformer le visage-test de manière à ce qu'il se présente dans des conditions moins difficiles et plus favorables.

3.4.4 Les occultations partielles

Le visage peut être partiellement masqué par des objets dans la scène, ou par le port d'accessoires tels que les lunettes de soleil, une écharpe, un chapeau, ... Les occultations peuvent être intentionnelles ou non. Dans le contexte de la vidéosurveillance, par exemple, il peut s'agir d'une volonté délibérée de contrecarrer la reconnaissance. La plupart des systèmes déployés doivent être non intrusifs, i. e. ne comptant pas 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 [MB98]. Leurs

résultats expérimentaux montrent que les performances des algorithmes testés sont relativement faibles dans ces conditions par rapport à l'utilisation de visages entièrement découverts. Les méthodes globales telles que les *Eigenfaces* (PCA) [TP91] et les *Fisherfaces* (LDA) [SW96] sont plus performantes que l'Analyse des Caractéristiques Locales (ACL) [PA96] dans le cas où les yeux sont occultés par des lunettes noires. Les taux de reconnaissance sont de l'ordre de 45% pour la LDA contre 10% seulement pour l'ACL. Par contre, pour des occultations de la partie inférieure du visage, l'Analyse des Caractéristiques Locales (ACL) donne de meilleurs résultats. Le comportement des méthodes des *Eigenfaces* et des *Fisherfaces*, qui sont basées sur l'étude de l'ensemble de l'image du visage modélisé sous la forme de vecteur (méthodes globales) est facilement interprétable. En effet, l'occultation d'une importante région du visage entraîne la modification des valeurs d'une grande partie du vecteur image 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 [Vis.fr] doit sans aucun doute accorder une forte importance à la région supérieure du visage. L'algorithme est 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.

3.4.5 Le vieillissement et le changement 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 pour déterminer l'impact de ces facteurs. Sur la base AR, où le délai entre les deux sessions de prises de vue est seulement de deux semaines, la baisse des taux de reconnaissance est estimée à 20%. Dans le rapport FRVT 2000 [BBP01], les effets du temps ont été mesurés à l'aide des vues *duplicate* de la base FERET : celles-ci sont comparées aux vues "fa" de la galerie, ce qui permet d'établir un taux de reconnaissance. Les taux fournis par les meilleurs algorithmes sont de 63% sur les vues "dup1", et 64% pour les vues "dup2", contre 58% et 52% pour l'évaluation de 1996. Notons que l'ensemble de test dup1 est formé par toutes les vues prise à des sessions ultérieures que les vues "fa" de la galerie. L'ensemble "dup2" est un sous ensemble de "dup1" formé par toutes les vues prises au moins 540 jours après celles qui forment la galerie.

Plus récemment, les systèmes de reconnaissance ont réalisé plus de progrès pour gérer au mieux le délai de temps entre deux prises de vue [Phi07]. Néanmoins, bien que l'intervalle de temps entre les vues "fa" et les vues "dup1" ne soit pas important, les systèmes ont du mal à reconnaître ces dernières. Cela provient certainement des changements dans les conditions de prises 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.

3.5 Les méthodes de reconnaissance automatique de visages

Plusieurs méthodes de reconnaissance de visage ont été proposées durant les trois dernières décennies. La reconnaissance du visage est devenue un challenge intéressant qui a attiré les chercheurs de différentes spécialités : psychologie, reconnaissance de forme, réseaux de

neurones, vision par ordinateur, infographie ... Suite à cela la littérature en reconnaissance du visage est vaste et diverse. Souvent un seul système implique des techniques motivées par différents principes. L'utilisation d'une mixture de techniques rend difficile la classification de ces systèmes suivant le type des techniques employées pour la représentation et la classification des caractéristiques. Nous nous sommes basés essentiellement sur les travaux de Zhao [Zha03], de Li [LJ04], de Visani [Vis05], et de Wechsler [Wec06] pour rédiger cet état de l'art. Pour avoir une catégorisation claire et de haut niveau des méthodes utilisées, nous avons classifié les travaux suivant l'approche de reconnaissance humaine suggérée par [Zha03] à savoir la caractérisation globale et/ou locale des images de visages :

1. *Méthodes globales.* Ces méthodes utilisent la totalité du visage comme vecteur d'entrée du système de reconnaissance. Une des représentations la plus utilisée pour coder le visage est l'*Eigenfaces* [KS90], qui est basée sur l'analyse en composantes principales.
2. *Méthodes locales.* Typiquement, dans ces méthodes, les caractéristiques locales telles que les yeux, le nez et la bouche sont extraites et leurs positions ainsi que leurs statistiques locales (géométriques et/ou d'apparences) sont fournies à un classifieur structurel.
3. *Méthodes hybrides.* Un système de reconnaissance utilise les caractéristiques locales et la région globale du visage pour le reconnaître comme le fait le système de perception humaine. Ces méthodes peuvent potentiellement offrir un meilleur taux de reconnaissance que les deux types de méthodes précédentes.

L'information globale et les caractéristiques locales sont cruciales pour la perception et l'identification des visages [WPB⁺98]. Les études suggèrent la possibilité de descriptions globales avant une perception plus fine basée sur les caractéristiques locales. Si les caractéristiques dominantes sont présentes, on peut ne pas utiliser les descriptions globales. Par exemple, dans les études de mémorisation de visage, on se concentre rapidement sur les caractéristiques « bizarroïdes » telles que les grandes oreilles, un nez tordu, ...

3.5.1 Les approches globales

Dans cette section nous allons passer en revue les principales techniques globales pour la représentation des visages en vue de leur reconnaissance. Nous développons principalement les techniques de projections à savoir les *Eigenfaces*, les *Fisherfaces* et l'Analyse en Composantes Indépendantes (*Independent Component Analysis - ICA*). Dans tous ces algorithmes de projection, la reconnaissance est établie en projetant l'image d'entrée \mathbf{x} dans un sous-espace par l'intermédiaire d'une matrice de projection P :

$$\mathbf{z} = P \mathbf{x} \quad (3.1)$$

Ensuite, on compare le vecteur \mathbf{z} des coefficients de projection de l'image d'entrée à tous les vecteurs préenregistrés, des individus enrôlés, pour déterminer la classe de la personne à identifier.

3.5.1.1 L'identification à base d'*Eigenfaces*

Utilisant initialement les méthodes Karhunen-Loeve (*KL*) et l'Analyse en Composantes Principales (*PCA*) pour la compression des visages en dimension réduite [KS90, SK87], les *Eigenfaces* ont été utilisés par la suite pour la représentation et l'identification de visage. La

formulation de cette technique de représentation des visages est développée dans la section 2.5.2 du deuxième chapitre.

Un avantage d'employer de telles représentations est leur sensibilité réduite au bruit. De bonnes performances sous le flou, l'occlusion partielle et les changements du fond ont été démontrés dans beaucoup de systèmes basés sur la PCA [Zha03]. Ceci est justifiable par le fait que l'aspect global des images reconstruites par PCA est meilleur que celui des images originales bruitées.

En utilisant une mesure probabiliste de similarité, au lieu de la simple distance euclidienne utilisée par [TP91], l'approche standard *Eigenfaces* a été étendue par Moghaddam et Pentland [MP97] à une approche bayésienne. Pratiquement, l'inconvénient principal d'une méthode bayésienne est la nécessité d'estimer, avec un nombre très limité d'échantillons par classe (individu), les distributions de probabilité de chaque classe dans un espace de grande dimension. Les auteurs ont évité ce problème en considérant simplement deux classes et en utilisant une mesure de similarité basée sur une analyse bayésienne entre les images. Deux classes mutuellement exclusives ont été définies représentant la variation intra personnelle entre les multiples images d'un même individu Ω_I , et la variation entre différents individus Ω_E . La classe de variation intra personnelle a été modélisée par le sous espace F obtenu par les k premières colonnes de la matrice orthogonale Φ . Alors que la classe de variation entre différents individus a été représentée par le sous-espace complémentaire \bar{F} . L'hypothèse établie estime que les données ont au maximum une dimensionnalité intrinsèque k et résident dans F alors que \bar{F} représente un bruit additif Gaussien. Supposant que les deux classes sont de distribution Gaussienne, les fonctions de probabilité $P(\Delta/\Omega_I)$ et $P(\Delta/\Omega_E)$ ont été estimées pour une différence $\Delta = I_1 - I_2$ entre les images I_1 et I_2 . Etant donné ces fonctions de probabilité et en utilisant la règle du maximum *a posteriori* MAP, la décision que deux images de visage correspondent au même individu sera prise si $P(\Delta/\Omega_I) > P(\Delta/\Omega_E)$. Moghaddam et Pentland [MP97] rapportent une amélioration des performances par cette technique probabiliste avec un appariement du plus proche voisin des vecteurs propres en utilisant un ensemble de données de visage comprenant la base de données FERET [PMRR00]. Cependant les dimensions M des sous-espaces Ω_I et Ω_E restent difficiles à estimer. Par exemple, $M_I = 10$ et $M_E = 30$ ont été utilisées pour des essais internes, alors que $M_I = M_E = 125$ étaient employées pour les tests de FERET. Sur la figure 3.1, les *Eigenfaces* duals séparément entraînés sur des échantillons de Ω_I et Ω_E sont présentés avec les *Eigenfaces* standard [MP97]. Tandis que les *Eigenfaces* extra personnel ressemblent plus aux *Eigenfaces* standard que les *Eigenfaces* intrapersonnel, les *Eigenfaces* intra personnel représentent des variations subtiles dues la plupart du temps à l'expression et à l'éclairage, suggérant qu'ils sont plus critiques pour l'identification.

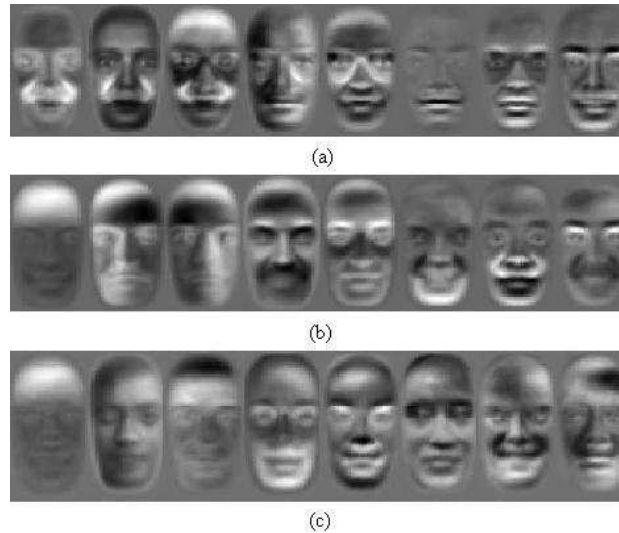


Figure 3.1 : Comparaison des *Eigenfaces* dual et des *Eigenfaces* standard: (a) interpersonnel, (b) extrapersonnel, (c) standard (Moghaddam et Pentland [MP97]).

3.5.1.2 L'identification à base de *Fisherfaces*

Plusieurs systèmes d'identification du visage par l'application de l'analyse discriminante linéaire (*Linear Discriminant Analysis - LDA*) ont été proposés par [SW96, BHK97, EC97, ZCK98]. L'apprentissage de la *LDA* est effectué par l'analyse de la matrice de dispersion [Fuk90]. Pour un problème à k classes (nombre de personnes dans la base de données), les matrices de dispersions inter et intra classes S_b et S_w sont calculées comme suit:

$$S_b = \sum_{i=1}^k \Pr(w_i) (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \quad (3.2)$$

$$S_w = \sum_{i=1}^k \Pr(w_i) C_i \quad (3.3)$$

où k est le nombre d'individus de la galerie pour construire l'espace de projection Φ et $\Pr(w_i)$, la probabilité à priori de la classe w_i , est généralement remplacée, dans la pratique, par $\frac{1}{M}$ avec l'hypothèse d'égalité des probabilités. S_w est la matrice de dispersion intra classe impliquant la matrice de covariance C_i des vecteurs x des différentes classes w_i avec leurs moyennes respectives \bar{x}_i : $C_i = E[(x - \bar{x}_i)(x - \bar{x}_i)^T]$. De même, S_b est la matrice de dispersion interclasse, représentant la variation de la moyenne de chaque classe \bar{x}_i avec la moyenne globale \bar{x} . Intuitivement, la *LDA* cherche les axes de projections maximisant la séparation linéaire des données. Par comparaison, notons que la *PCA* cherche à maximiser la variance le long des axes représentant les données. On considère qu'un sous espace est discriminant s'il minimise par projection les variations à l'intérieur des classes, tout en maximisant les variations entre classes.

Quatre critères ont été utilisés pour mesurer le pouvoir discriminant d'un sous espace dont le critère de Fisher est le plus utilisé dans le domaine de la reconnaissance de forme [Vis05]. Pour choisir le sous-espace linéaire Φ optimal, le critère de Fisher maximise le rapport du

déterminant de la matrice de dispersion interclasse des échantillons projetés par le déterminant de la matrice de dispersion intra-classe :

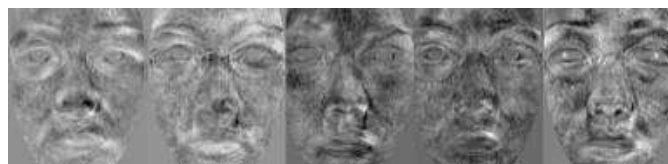
$$\Phi_{opt} = \arg \max_{\Phi} \frac{|\Phi^T S_b \Phi|}{|\Phi^T S_w \Phi|} \quad (3.4)$$

La matrice de projection optimale Φ qui maximise ce ratio peut être obtenue en résolvant le problème suivant des valeurs propres:

$$S_b \Phi = S_w \Phi \Lambda_{\Phi} \quad (3.5)$$

Où Λ_{Φ} est la matrice diagonale des valeurs propres des axes de l'espace Φ . Sous l'hypothèse que la matrice S_w est inversible, les colonnes de la matrice Φ sont constituées des vecteurs propres de la matrice $S_w^{-1} S_b$, associés aux plus grandes valeurs propres.

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 ($n \gg N$). La matrice S_w est alors non inversible et on ne peut pas déterminer directement Φ : c'est le problème de la singularité. Une solution à ce problème, proposée par Belhumeur [BHK97] et Swets et Weng [SW96] est d'effectuer une *PCA* en amont de la *LDA*. 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 matrices de projection Φ , dont des exemples sont fournis en figure 3.2, sont appelés *Fisherfaces* par Belhumeur [BHK97] et *Most Discriminant eigenfeatures* dans [SW96]. Nous les désignons sous le nom de *Fisherfaces* dans la suite. Visuellement, les *Fisherfaces* sont moins représentatifs de la structure des visages que les *Eigenfaces* puisque le critère à maximiser n'est pas lié à la qualité de représentation, mais à la séparabilité. Au contraire des *Eigenfaces* illustrés à la figure 3.1, on constate que les *Fisherfaces* semblent moins représentatifs des variations d'éclairages. En effet, celles-ci sont en grandes partie expliquées par la variance intra-classe, que le critère de Fisher tend à minimiser.



Figures 3.2 : Exemple de cinq premiers *Fisherfaces*, associés aux plus grandes valeurs propres. Extrait de [Vis05].

3.5.1.3 L'identification à base de l'Analyse en Composantes Indépendantes

Basé sur l'argument qu'une grande partie de l'information faciale est contenue dans les statistiques d'ordre élevé, Bartlett et al. ont proposé dans [BLS98] d'utiliser l'Analyse en Composantes Indépendantes (en français *ACI*; en anglais: *Independent Component Analysis* : *ICA*) pour extraire des caractéristiques pour l'identification de visage. Au lieu de simplement décorréler les données comme le font la *PCA* et la *LDA*, l'*ICA* cherche à les rendre statistiquement indépendant. La méthode *ICA* proposé dans [BLS98] repose sur une étape préliminaire de l'analyse en composantes principales *PCA*. Deux architectures ont été

proposées pour l'identification de visage : la première architecture vise à obtenir des vecteurs *Eigenfaces* qui soient statistiquement indépendant deux à deux. La deuxième architecture, quant à elle, cherche à rendre les coefficients de projection (variables) mutuellement statistiquement indépendant.

Les vecteurs de projection de l'ICA, à l'instar des *Eigenfaces* pour la PCA et des *Fisherfaces* pour la LDA, sont illustrés en figure 3.4. l'ICA n'engendre pas de réduction de dimension. Par conséquent, les vecteurs propres sont en même nombre que la dimensionnalité choisie pour la PCA préliminaire. Il est à 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.



Figure 3.3: Extrait de [DBBB03]. Les six premiers axes de l'espace de visage obtenus (a) grâce à l'architecture I et (b) avec l'architecture II.

Dans [BLS98], la PCA est employée tout d'abord pour produire un espace de taille réduite par rapport à celui des images de taille (60×50). L'ICA est exécutée sur les 200 premiers vecteurs propres dans la première architecture, et sur les 200 premiers coefficients de projection de la PCA dans la deuxième architecture. Les auteurs ont rapporté une amélioration des performances des deux architectures par rapport aux *Eigenfaces* standard avec un ensemble de 425 individus de la base FERET.

Dans la littérature, ces résultats sont contradictoires avec ceux obtenus par Drapper et al. [DBBB03] qui montrent une différence très significative dans les performances des deux algorithmes, à l'avantage de l'architecture II. Selon cette référence, seul le modèle issu de la seconde architecture est plus efficace que la PCA. Néanmoins, la plupart des résultats expérimentaux laissent à penser que l'ICA n'apporte pas d'amélioration significative sur la PCA [Mog02], voire qu'elle engendre une dégradation des performances d'après [BDBS02] et comme nous l'avons remarqué sur nos échantillons de données.

Une problématique s'est posée pour ces trois dernières méthodes de projection sur les bases de données évolutives et notamment les grandes bases de données où on a souvent besoin d'ajouter de nouvelles personnes ou d'en supprimer quelques unes. En effet, vue la difficulté de mettre constamment à jours les bases de projection de la PCA et de la LDA, il est intéressant d'examiner les possibilités adaptatives de ces méthodes en fixant leurs bases de projection quand des images de nouvelles classes sont ajoutées. Zhao et al. [ZCK98] ont fixé la dimensionnalité du sous-espace de la PCA à 300 en faisant l'apprentissage avec un grand nombre d'échantillons. Un ensemble de 4056 vues de face construites à partir des 1078 images originales de FERET de 444 individus a été employé en ajoutant du bruit et l'effet miroir. Penev et Sirovich ont avancé dans [PS00] que la dimensionnalité du sous-espace global de visages doit être de l'ordre de 400 pour de grandes bases de données de 5.000 images. Dans [Zha99], Zhao essaye de régulariser l'apprentissage de la LDA en modifiant la matrice de variation intra classe S_w à $S_w + \delta I$ où δI est un nombre positif relativement petit. Faire ceci résout aussi le problème numérique de la singularité de S_w . Dans le cas où seulement un échantillon par classe est disponible, cette régularisation transforme le problème de la LDA en un problème standard de PCA avec S_b étant la matrice C de covariance.

3.5.1.4 L'identification à base de moments de Zernike

Les moments de Zernike constituent une famille de fonctions orthogonales ce qui permet la génération de descripteurs non redondants. En plus, ils présentent des propriétés d'invariance en translation, en échelle et en rotation [BW54]. Plusieurs formulations et variantes des moments de Zernike ont été étudiées dans la description et la reconnaissance de visages.

Ono [Ono03] a développé les moments de Zernike circulaires pour extraire les signatures faciales. La comparaison de ses résultats avec ceux des moments classiques établis par les équations 2.4 et 2.5 a enregistré une légère amélioration du taux de reconnaissance de 95,87% à 96,18% sur une base de visage non communiquée de 27 personnes.

Dans [HAF03, Ali05, KFG08], les auteurs ont utilisé les Pseudos Moments de Zernike (*PZM*) pour extraire les attributs. Pour un ordre donné, les pseudos moments de Zernike extraient deux fois plus d'attributs que les moments classiques. Du coup, les pseudos moments de Zernike sont considérés plus performants que les moments classiques bien qu'aucune comparaison directe entre ces deux algorithmes n'ait été présentée. Pang et al. [PTN06] ont effectué récemment une comparaison entre les pseudos moments de Zernike et les méthodes des *Eigenfaces* et *Fisherfaces*. Les résultats illustrés dans le tableau 3.1 sont comparables avec un léger avantage pour les *Fisherfaces*. Il est à noter que ces expérimentations ont été menées sur la base Essex qui contient 72 personnes et 20 poses pour chacun. Dix images ont été utilisées pour l'apprentissage de la *PCA* et la *LDA* et les dix autres pour les tests.

méthode	PZM	PCA	LDA
Taux d'identification	95,80%	95,89%	96,30%

Tableau 3.1 : Comparaison des pseudos moments de Zernike avec les *Eigenfaces* et les *Fisherfaces*. Extrait de [PTN06].

3.5.1.5 Autres représentations

D'autres techniques basées sur la construction d'un sous espace non linéaire ont été introduites dans le but de parvenir à représenter plus précisément les données faciales. Typiquement, si les conditions d'éclairage changent drastiquement, les techniques non linéaires sont réputées plus performantes que les méthodes linéaires.

La Poursuite d'évolution (PE). Une représentation adaptative basée sur la poursuite d'évolution (*evolution pursuit*) et son application à la reconnaissance faciale a été présentée dans [LW00]. Par analogie aux méthodes de projections, la poursuite d'évolution (*PE*) cherche à apprendre une base optimale pour des applications de compression de données et de classification des formes. Afin d'augmenter les capacités de généralisation de la méthode, un équilibre est cherché entre la minimisation du risque empirique produit pendant l'apprentissage et le rétrécissement de l'intervalle de confiance pour réduire le risque en examinant les données invisibles [Vap95]. Pour cette finalité, la *PE* implémente des stratégies d'algorithmes génétiques (GA) pour rechercher l'espace des solutions possibles et déterminer la base optimale. La *PE* commence par projeter les données originales par la *PCA* dans un espace de dimension réduite. Des rotations aléatoires des vecteurs de la base dans cet espace sont alors recherchées par les algorithmes génétiques où l'évolution est conduite par une fonction définie en termes de précision des performances (risque empirique) et de séparation de classes (intervalle de confiance). La faisabilité de cette méthode a été démontrée pour

l'identification de visage, où le grand nombre des bases possibles exige un algorithme de recherche très gourmand. L'identification du visage a été réalisée sur 1107 images frontales de visage de 369 sujets de la base FERET, avec trois images frontales pour chaque sujet, deux pour l'apprentissage et la troisième pour le test. Les auteurs font état d'une amélioration des performances de l'identification par rapport aux *Eigenfaces* [TP91], et une meilleure capacité de généralisation que les *Fisherfaces* [BHK97].

Les courbes Principales. La technique des Courbes Principales [Tib92] est un modèle de régression non linéaire des données. L'un des algorithmes de mise en œuvre les plus simples est basée sur la construction d'une *PCA* non linéaire par le biais d'un réseau de neurones multicouches auto-associatif [CF90, Kra91]. Cette technique permet de trouver un sous espace optimal au sens de la reconstruction des données initiales et est donc, tout comme la *PCA*, plus adapté à la compression des données qu'à leur classification.

Utilisation d'une fonction de noyau. L'utilisation d'une fonction de noyau, conjointement avec un classifieur linéaire, a été conçue pour séparer correctement les classes dans l'espace initial des données. La technique a été publiée initialement en 1964 par Aizerman [ABR64] et n'a été appliquée ensuite qu'avec Boser [BGV92]. Elle consiste à transformer l'espace de représentation des données d'entrées en un espace K de plus grande dimension, où le classifieur linéaire est alors utilisé. Ainsi, la discrimination linéaire dans l'espace de grande dimension est équivalente à une discrimination non-linéaire dans l'espace d'origine. Notons qu'il n'est pas nécessaire de définir précisément l'espace K mais qu'il suffit de choisir une fonction de noyau adaptée au problème. Des versions à noyau ont ainsi été proposées pour la *PCA* [SSM98, Yan02, LLT08], la *LDA* [MRW⁺99], et l'*ICA* [BJ02]. Cependant, le choix du type de fonction de noyau à utiliser, ainsi que ses paramètres, reste un problème difficile [GAP+02].

Les modèles Actifs d'Apparence (Active Appearance Models - *AAM*) est une technique d'extraction de signatures caractérisant à la fois la forme et la texture d'un visage. Les *MAA* est une technique très utilisée dans le contexte de la reconnaissance faciale, ainsi que dans la détection, le suivi des visages et l'analyse et la reconnaissance d'expression faciale [LTC95, CET01, KAG03, CC04].

De nombreuses autres méthodes ont été utilisées pour la reconnaissance de visages tels que les transformations d'ondelettes et notamment les ondelettes de Gabor, les méthodes basées directement sur les réseaux de neurones [Law97, Tan05], les Machines à Vecteurs de Support [Phi98] ...

3.5.2 Les approches d'identification basées sur les caractéristiques locales et hybrides

Dans cette section, nous présentons les principales approches locales basées sur l'étude de caractéristiques extraites de différentes régions des visages. Nous exposons également les techniques hybrides, en ce sens qu'elles utilisent conjointement des caractéristiques globales et locales des visages.

Plusieurs méthodes dans la catégorie d'appariement structurel de caractéristiques locales ont été proposées, incluant plusieurs anciennes méthodes basées sur la géométrie des caractéristiques locales [Ke170, Kan77] aussi bien que les méthodes des champs de Markov cachés HMM à une dimension [SY94] et à deux dimensions [CL08]. Un système des plus

réussi est celui de l'appariement élastique de groupe de graphes (*Elastic Bunch Graph Matching : EBG*M) [Oka98, Wis97], qui est basé sur l'architecture des liens dynamiques (*Dynamic Link Architecture : DLA*) [Lad93].

3.5.2.1 Les approches géométriques

Les méthodes géométriques furent les premières à être utilisées pour la reconnaissance automatique des visages. Elles sont basées sur l'extraction des positions relatives des éléments constituant le visage à savoir le nez, la bouche et les yeux. Ainsi, la plupart des approches géométriques utilisent les coins et les centres de la bouche, des yeux et du nez comme des points d'intérêts. Au début des années 90, Brunelli et Poggio [BP93] ont décrit un système de reconnaissance faciale qui extrait automatiquement 35 caractéristiques géométriques du visage. La reconnaissance est effectuée par la classification bayésienne. Le coût de stockage des techniques géométriques est très bas comparé à celui des autres techniques. Toutefois les approches de reconnaissance purement géométriques ne suffisent pas à elles seules à représenter un visage par insuffisance d'informations discriminatoires. De plus, elles nécessitent une très grande précision de détection des divers éléments faciaux et reste non robuste à des changements d'expressions faciales ou de pose de la tête.

Takacs a utilisé pour la première fois, en 1998, les cartes de contours calculées par le biais du filtre de Sobel pour la reconnaissance de visage [Tak98]. Les cartes de contours présentent notamment l'avantage d'être robustes à des changements d'éclairage dans les images. L'approche consiste à comparer les images de visages par une mesure de similarité entre les cartes de contours binaires sans pour autant nécessiter une 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 des 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 (LEM)*. Les auteurs ont rapproché cette technique des études biologiques qui ont montré la capacité du cerveau humain à reconnaître un visage depuis son dessin ou sa caricature. Les expérimentations réalisées sur la base AR (voir chapitre suivant) montrent que la technique LEM est plus efficace que celle de Takacs [Tak98]. Elle donne également des résultats de classification significativement meilleurs que les *Eigenfaces* en présence de changements d'éclairage. En revanche, elle est moins robuste aux variations d'expression faciale et d'angle de prise de vue.

3.5.2.2 Les techniques modulaires

Les méthodes modulaires présentées dans cette section sont classées parmi les approches hybrides d'identification. Elles reposent sur des techniques globales 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. Ainsi on espère apporter une certaine robustesse, essentiellement vis-à-vis des changements d'expression faciale et des occultations partielles.

Pentland *et al.* ont introduit en 1994 l'approche modulaire des *Eigenfaces* [PMS94] qui utilise cette représentation sur les caractéristiques globales et locales des visages. Le concept des *Eigenfaces* a été étendu aux *eigenfeatures*, sur les régions faciales telles que les yeux et le nez. Les mesures de similarités obtenues sur chacune des régions sont agrégées. La bouche étant

trop sensible à des changements d'expression faciale, sa prise en compte engendrait une baisse des taux de reconnaissance. Ainsi, les auteurs ont montré que cette technique est plus efficace que les techniques globales et celles strictement locales. Dans [Gue08], Guerfi a développé cette technique en combinant les différents modules des yeux, du nez et de la bouche avec le modèle global du visage. Elle a développé notamment une pondération adaptative de chaque module suivant ses performances à travers son taux *EER*.

En 2001, Heisele [HHP01] a introduit une technique modulaire en utilisant les Machines à Vecteurs de Support (SVM). Quatorze régions faciales sont détectées et dix d'entre elles, les moins fortement chevauchées, sont extraites pour la reconnaissance faciale. Chaque région est normalisée puis ses niveaux de gris sont combinés avec ceux des autres régions pour construire un vecteur caractéristique représentant le visage. La motivation de l'usage des composantes est que les changements de la pose de la tête mènent principalement aux changements des positions des composantes faciales. Cependant, un inconvénient important du système est qu'il a besoin d'un grand nombre d'images d'apprentissage, prises sous différents points de vue et dans différentes conditions d'éclairage. Pour surmonter ce problème, Huang et al. [HHB03] ont utilisé le modèle déformable du visage 3D pour produire des images synthétiques arbitraires sous des poses et des éclairages variables. Trois images (frontale, 3/4 profil, profil) d'une personne sont utilisées pour calculer le modèle du visage 3D. Une fois que le modèle 3D est construit, des images synthétiques sont produites pour l'apprentissage du détecteur et du classificateur. Huang et al. ont utilisé seulement neuf régions, parmi les quatorze détectées, pour la reconnaissance. En outre, la région du visage a été ajoutée aux neuf composantes pour former un vecteur de caractéristiques qui a été entraîné par les SVM. Cette représentation modulaire a donné de meilleurs résultats d'authentification que la même représentation sur le visage global.

Dans [PG05], Price et Gee ont introduit une technique modulaire basée sur une variante de l'Analyse Discriminante Linéaire (*LDA*). Le visage global a été combiné avec une bande faciale, de même largeur, s'étalant du front jusqu'au dessous du nez et une bande faciale contenant les yeux. Les résultats montrent que cette approche est plus performante que les techniques des *Eigenfaces* et des *Fisherfaces*, et notamment plus robuste aux changements d'éclairage du visage, d'expression faciale ainsi qu'aux occultations partielles.

3.5.2.3 L'Analyse des caractéristiques locale

L'Analyse des caractéristiques Locale (*Local Feature Analysis : LFA*) est une méthode intéressante d'analyse de caractéristiques inspirée des propriétés biologiques de la rétine humaine où une petite fraction des cellules réceptrices est activée correspondant aux objets/signaux naturels qui sont statistiquement redondants [PA96]. Pour explorer cette redondance, la *LFA* est employée pour extraire les caractéristiques locales topographiques à partir des modes globaux de la *PCA*. À la différence des noyaux de la *PCA* qui ne contiennent aucune information topographique (leurs supports s'étendent sur la grille entière des images), les noyaux de la *LFA* $K(\mathbf{x}_i, \mathbf{y})$ sont définis sur des grilles choisies \mathbf{x}_i en un support local³. Ces grilles, ainsi que leurs noyaux associés, sont illustrées sur la figure 3.4.

La recherche du meilleur ensemble topographique de grilles distribuées d'une façon éparse a été basée sur l'erreur de reconstruction. Selon Penev et Atick [PA96], en utilisant le même nombre de noyaux, la qualité perceptuelle de reconstruction de la *LFA* basée sur l'ensemble

³ Ces noyaux indexés par les grilles \mathbf{x}_i sont similaires aux noyaux de l'ICA dans la première architecture ICA [BLS98].

optimal de grilles est meilleure que celle de la PCA. Néanmoins, aucun résultat de reconnaissance n'a été publié dans les articles de Penev et Atick [PA96] et de Penev et Sirovich [PS00] pour évaluer cette technique de représentation. Selon [Vis05], cette technique souffre de deux inconvénients majeurs : la taille des signatures des visages qui est beaucoup plus importante qu'avec la méthode des *Eigenfaces* et le coût de la procédure d'optimisation itérative. De plus, les caractéristiques sont choisies de manière à être les plus représentatives possible des images de visages, mais non dans un but de séparation des classes. Néanmoins, le groupe Visionic [Vis.fr] revendique l'utilisation de la LFA dans le logiciel commercial de reconnaissance de visages FaceIt.

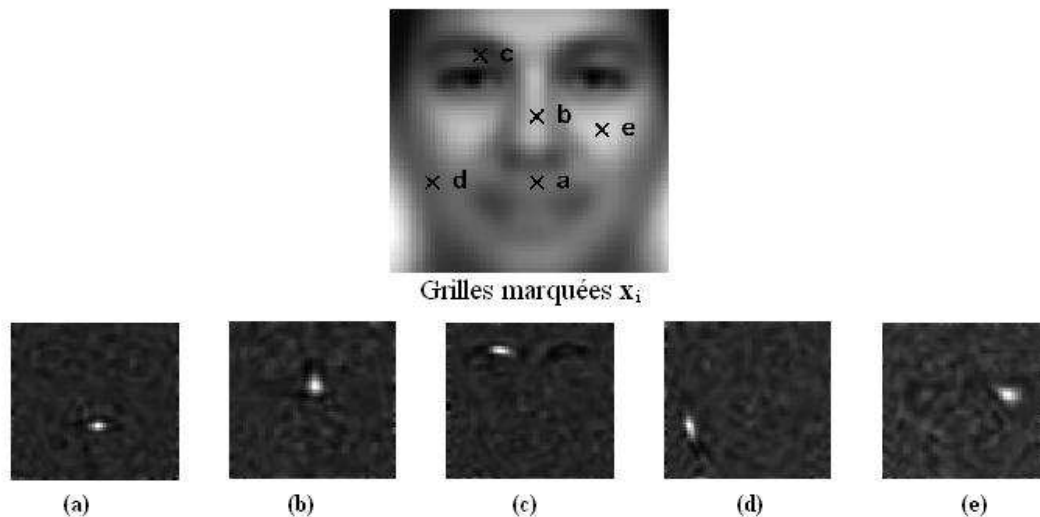


Figure 3.4. L'image du visage donne les positions des centres des différentes grilles locales. Les images (a) à (e) montrent les noyaux locaux associés à ces grilles. Extrait de Penev et Atick [PA96].

3.5.2.4 Les techniques d'identification basées sur les graphes

Les techniques d'identification présentées dans cette partie sont basées sur la mise en correspondance de graphes [Con04]. Nous nous intéressons particulièrement aux approches appelées *Elastic Graph Matching* et *Elastic Bunch Graph Matching* [AMM08]. Ces méthodes sont basées sur la méthodologie d'*Architecture de Lien Dynamique (DLA)*. Cette dernière est étroitement liée à la théorie des réseaux de neurones. Elle permet de structurer, par le biais de graphes, des neurones caractérisant des propriétés locales des visages. Une représentation typique des caractéristiques locales s'effectue en affectant des coefficients d'ondelettes pour différentes échelles et rotations (appelées *jets* par Okada et al. [Oka98]). Ces coefficients d'ondelettes, localement estimés, sont robustes au changement d'éclairage, à la translation, à la déformation, à la rotation, et aux facteurs d'échelles. La fonction de base de Gabor 2D et sa transformée de Fourier sont :

$$g(x, y : u_0, v_0) = \exp(-[x^2/2\sigma_x^2 + y^2/2\sigma_y^2] + 2\pi i [u_0x + v_0y]) \quad (3.6)$$

$$G(u, v) = \exp(-2\pi^2 (\sigma_x^2 (u - u_0)^2 + \sigma_y^2 (v - v_0)^2)) \quad (3.7)$$

où σ_x et σ_y représentent les largeurs de la gaussienne et (u_0, v_0) les fréquences de la sinusoïde.

La technique *Elastic Graph Matching (EGM)* a été initialement introduite par Lades et al. en 1993 [Lad93]. Les images sont représentées par un graphe dont les nœuds correspondent aux points fiduciels (tels que les yeux, le bout du nez, ...), et les arrêtes représentent les vecteurs de distance 2D (cf. figure 3.5). Chaque nœud est enregistré avec son jet formé de 40 coefficients complexes des ondelettes de Gabor, incluant la phase et l'amplitude. Les coefficients des Ondelettes sont extraits en utilisant un banc de filtres de Gabor ayant 5 fréquences spatiales et 8 orientations différentes. L'architecture de lien dynamique (*DLA*) modélise ce graphe de nœud comme un réseau de neurones structuré en assimilant les nœuds du graphe à des neurones groupés et en adaptant les connexions entre ces derniers suivant les coefficients de Gabor enregistrés. L'identification d'une nouvelle image se fait en transformant l'image en grille de jets, et par appariement de tous les modèles de graphes enregistrés.

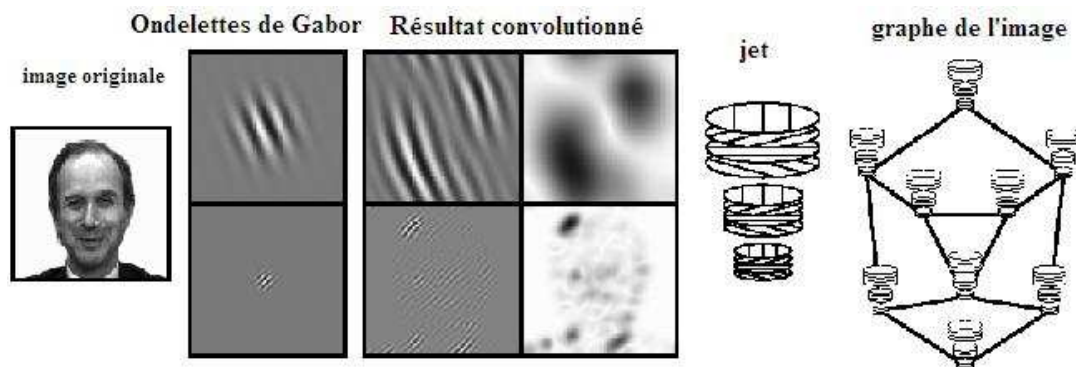


Figure 3.5. Représentation élastique de graphe. Extrait de Wiskott et al. [Wis97].

L'architecture de la *DLA* a été, plus tard, étendue à l'appariement élastique de groupe de graphe [Wis97] (cf. figure 3.6). Ceci est semblable au graphe décrit ci-dessus, mais au lieu d'attacher seulement un seul jet à chaque nœud, les auteurs ont attaché un ensemble de jets, chacun dérivé d'une image différente du visage. 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 3.6). 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.

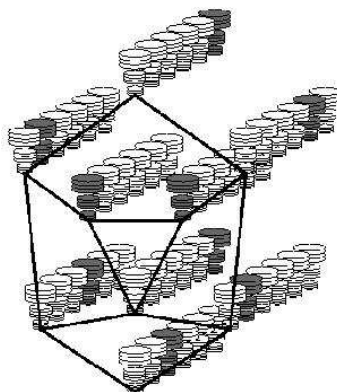


Figure 3.6. Un Face Bunch Graph (FBG) qui vise à représenter l'ensemble des états possibles, pour chaque nœud. Extrait de Wiskott et al. [Wis97].

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 distance entre l'image requête et une image connue est définie comme étant la meilleure mise en correspondance entre les vecteurs de caractéristiques des deux images parmi les solutions possibles. Etant donné la combinatoire du modèle, le nombre de solutions possibles est très important. C'est pourquoi une fonction de similarité, incluant la somme des coûts locaux de mise en correspondance et le coût de la déformation globale du modèle, a été proposée.

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. Dans [PMRR00], les résultats de l'évaluation FERET comparant les taux d'identification des méthodes *Eigenfaces*, *Fisherfaces* et *EBGM* sont communiqués (voir tableau 3.2). Deux ensembles de test différents de la base FERET acquis dans des conditions similaires de la galerie ont été utilisés pour évaluer l'identification sur une base de 1196 personnes. Le premier ensemble de test, appelé *FB*, contient des images acquises le même jour que l'image correspondante de la galerie. Le deuxième ensemble de test, appelé *duplicateI*, contient des images collectés dans des sessions ultérieures à celles de la galerie. Les deux systèmes les plus performants sont l'*EBGM* et les *Fisherfaces*. La méthode *EBGM* fournit de meilleurs résultats que les *Fisherfaces* sur les vues duplicate. Pour les vues 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 la *LDA*.

	FB	duplicateI
<i>Eigenfaces</i>	80%	41%
<i>Fisherfaces</i>	96%	48%
<i>EBGM</i>	95%	59%

Tableau 3.2 : Comparaison des performances des méthodes sur la base de visages FERET. Extrait de [PMRR00].

3.6 Conclusion

Ce chapitre nous a permis de tirer un certain nombre d'enseignements du processus de reconnaissance humaine des visages. Après le développement des principales difficultés inhérentes à la reconnaissance automatique des visages, nous avons présenté les principales techniques 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 locales des visages. Les techniques hybrides, utilisent conjointement ces deux types de modélisation à l'instar du fonctionnement du système visuel humain. Ces dernières

techniques sont plus robustes et performante à des changements d'apparence du visage que les techniques globales. Néanmoins, elles sont généralement beaucoup plus complexes et coûteuses en temps de calcul et moins stables en terme de reconnaissance.

Parmi les techniques globales, on compte notamment les méthodes factorielles basées sur la projection statistique, appelées aussi méthodes 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 ou de séparabilité de données. Pour optimiser le premier critère, on utilise essentiellement la *PCA* (technique des *Eigenfaces*) ou l'*ICA* tandis que, pour le second critère, la *LDA* (technique des *Fisherfaces*) est plus appropriée. Ces techniques sont caractérisées par un apprentissage assez rapide, un faible nombre de paramètres à ajuster et de très bonnes performances.

Nous avons développé, dans un premier temps, la méthode des *Eigenfaces* et celle des *Fisherfaces* pour participer aux évaluations menées dans le cadre du projet IV². Ces deux méthodes de représentation des caractéristiques faciales sont ensuite utilisées dans le cadre de notre approche d'identification. En plus de ces deux méthodes de représentation factorielles, nous avons utilisé les moments de Zernike qui extraient des caractéristiques géométriques à partir des images faciales. Ces caractéristiques transcrivent un autre type d'information des images de visage que les méthodes *Eigenfaces* et *Fisherfaces*. Ceci nous permettra notamment d'enrichir autant que possible les signatures représentatives du visage afin de relever à travers une signature certaines ambiguïtés ou défaillances causées par une autre signature.

Nous présentons, dans le chapitre suivant, les bases de données faciales utilisées dans notre système d'identification. Nous détaillons particulièrement le développement de la base IV² à la constitution de laquelle nous avons contribué.

Chapitre IV

Bases de données utilisées et corpus d'évaluation.

4.1 Introduction

Plusieurs bases de données contenant des informations qui permettent l'évaluation des systèmes de reconnaissance de visages sont disponibles pour la recherche et le développement de solutions biométriques efficaces. Toutefois, le nombre et les variétés croissantes des algorithmes de reconnaissance de visages proposés ces dernières années, ont imposé de définir des protocoles expérimentaux pour leur évaluation. La principale motivation de développer de tels protocoles est d'uniformiser et de standardiser les tests de performances des algorithmes de reconnaissance ce qui permettra leurs éventuelles comparaisons. Ainsi, la série d'évaluations FERET [PMRR00] menées par le «National Institute of Standards and Technology (NIST)» américain a permis la comparaison entre neuf systèmes de reconnaissance de visages proposés par des institutions et des entreprises dans le contexte de l'identification et de la vérification. D'autres évaluations ont été effectuées par la suite, notamment celles du «Face Recognition Vendor Test (FRVT)» [BBP01, PGM⁺03, Phi07]. Mais, l'évaluation des algorithmes de reconnaissance biométriques et spécialement de visages reste un problème difficile vu la multiplicité des situations rencontrées en cas réel. En effet, les différentes sources de variabilité relatives à l'angle et à l'environnement de prise de vue d'un objet solide sont innombrables. De plus, un visage humain constitue un objet déformable pouvant simuler une infinité de gestes et d'expressions faciales non quantifié ni mesurable. D'autre part, la mesure des performances doit se faire selon une méthodologie prenant en compte le coût des erreurs, en fonction des spécifications de l'application visée et de son niveau de sécurité.

Par conséquent, nous avons vu l'essor de plusieurs bases de données pour l'évaluation de la reconnaissance faciale pendant cette dernière décennie. Chacune de ces bases dispose d'attributs différents suivant les conditions d'acquisition des images de visages (changements d'éclairage, de pose, d'expression faciale), et le nombre de sessions d'acquisition. Les performances des algorithmes sont très dépendantes des variabilités incluses et du nombre de personnes enregistrées dans ces bases. Nous récapitulons dans la section 4.3 les principales caractéristiques de chacune de ces bases.

Nous développons dans l'annexe A une nouvelle base multimodale développée dans le cadre du projet national IV². Ce projet a été lancé par le ministère de la recherche et de l'industrie et le ministère de la défense dans le cadre de l'appel d'offre « Techno Vision ». Il a pour but de créer des ressources et les conditions d'une évaluation à l'échelle internationale de différents systèmes de reconnaissance biométrique liés à l'information du visage, de l'iris et de la voix. J'ai activement participé à toutes les phases de ce projet élaboré en partenariat avec d'autres

acteurs académiques et industriels français en biométrie⁴. Nous avons participé notamment à la conception, à la réalisation et à la maintenance de la cabine d'acquisition biométrique. Ensuite, nous avons défini le protocole d'acquisition et effectué la collecte des données sur trois sites universitaires. Nous avons également participé, avec cinq autres partenaires du projet, à la première série d'évaluations sur les données IV².

4.2 La base IV²

A l'issue du projet IV², nous avons acquis l'ensemble des données présentées dans le tableau 4.1.

<i>Nombre d'identités différentes</i>	315
<i>Nombre de personnes ayant fait une session unique</i>	219
<i>Nombre de personnes ayant fait deux sessions</i>	77
<i>Nombre de personnes ayant fait trois sessions</i>	19
<i>Nombre total de répertoires</i>	430

Tableau 4.1: Ensemble des données acquises lors de la campagne IV²

Ce sont donc plus de 300 personnes différentes qui sont passées dans le photomaton biométrique. Parmi ces individus, 77 ont accepté de venir une deuxième fois et même une petite vingtaine une troisième fois.

C'est ainsi que trois bases de données complémentaires sont désormais disponibles :

- La première correspond aux acquisitions des données d'une cinquantaine de personnes. Ces acquisitions ont été faites à la mise en route de la cabine, afin de valider son fonctionnement et d'affiner les procédures d'acquisition. Elles ne respectent donc pas scrupuleusement les protocoles décrits plus haut, mais peuvent être utilisées en tant que données de développement.
- La seconde correspond aux acquisitions réalisées sur une trentaine de personnes avec deux scanners 3D. Le scanner à laser MINOLTA décrit plus haut et le scanner GEOMETRIX, utilisé par notre partenaire EURECOM. Ce deuxième scanner utilise le principe de la stéréovision pour acquérir deux images d'une même personne et permettre la réalisation de son modèle 3D.
- La troisième, nommée base évaluation, comprend l'ensemble des données préparées pour les évaluations finales, qui sont détaillées au paragraphe 4.2 de ce document, ainsi que les annotations accompagnant les images et séquences dans le cadre de ces évaluations et qui sont présentées dans le prochain paragraphe.

A travers ces trois bases, ce sont donc quasiment un Tera octets de données qui sont disponibles pour les utilisateurs potentiels dans le cadre de campagnes d'évaluations.

⁴ Avec le laboratoire IBISC de l'Université d'Evry qui a coordonné le projet IV², les différents partenaires de ce projet sont : GET – INT, LIRIS - Ecole Centrale de Lyon, GET – ENST, IMEDIA – INRIA, Institut EURECOM, Let It Wave, THALES (TRT), URATEK

4.3 Les évaluations menées dans le cadre du projet IV²

Des compétitions internationales sont organisées depuis plusieurs années pour évaluer les systèmes commerciaux, ou les algorithmes de reconnaissance biométrique (FRVT, FRGC, ICE). De plus, des organismes spécialisés entreprennent maintenant des évaluations de systèmes commercialisés: Comparative Biometric Testing (CBT2006), par International Biometric Group *IBG*, Independent Testing of Iris Recognition Technology (ITIRT2005), également par *IBG*.

Dans ce contexte, nous avons décidé de commencer à exploiter la grande variété des données de la base IV², pour ouvrir la voie à des évaluations plus spécifiques.

4.3.1 La phase de développement

Pour permettre à tous les participants d'adapter leurs algorithmes au formalisme des données et des résultats proposés par l'évaluateur (THALES), une phase préliminaire d'évaluation a été réalisée en utilisant la première vague de données acquises et désignées précédemment par base de développement. Dans ce cadre de « réglage des algorithmes », des évaluations ont été réalisées, entre les partenaires du projet, sur le calcul des positions des yeux, et sur la vérification d'identité à partir de visages 2D, 3D et d'iris. Succinctement, 2000 images ont été utilisées pour les tests sur l'iris, 500 images issues du caméscope pour les tests sur le visage 2D, 285 prises de vues par le scanner 3D pour les tests sur le visage 3D et 510 images issues du caméscope pour les tests sur la position des yeux. Pour ce dernier test, la position du centre des yeux a été doublement positionnée manuellement afin de proposer une vérité terrain servant de référence à ces 510 images. Il est à noter que cette dernière évaluation sur la position des yeux n'a pas été reconduite dans la deuxième phase et cela pour deux motifs. Premièrement les algorithmes évalués se situent en amont de la phase de vérification ou d'identification biométrique et correspondent à une phase préparatoire, parfois nécessaire à certains algorithmes de reconnaissance biométrique. Deuxièmement, pour ce type d'évaluation, le recours à ce nombre important d'images, plus de 500, permet de penser que les résultats obtenus présentent une vérité statistique significative. Les résultats présentés dans cette phase de développement nous ont permis de valider l'utilisation d'un algorithme automatique de détection de la position des yeux pour annoter les images utilisées dans la phase finale des évaluations.

4.3.2 Conception et mise en œuvre des évaluations

L'objectif de ce paragraphe est de présenter de manière succincte les protocoles que nous avons retenus et les évaluations menées. Le détail de toutes ces évaluations ainsi que des analyses précises pour les différentes modalités sont disponibles dans le rapport du projet [CH07].

Les protocoles d'évaluation ont été mis au point au cours de plusieurs réunions rassemblant les partenaires du projet. Les contraintes qui ont guidé les choix sont les suivantes:

- évaluation technologique d'algorithmes et non de scénarios ou de systèmes
- recherche d'une homogénéité dans les méthodes et les résultats des diverses modalités
- possibilité pour les équipes de recherche de faire le test sur leurs équipements, et de transmettre seulement des fichiers de résultats.

Plusieurs évaluations ont été menées dans le cadre du projet sur l'iris, les visages 2D, les visages 3D, les visages parlant ainsi que des tests multimodaux regroupant plusieurs de ces modalités. Nous rapportons ici, les évaluations menées sur les visages 2D, auxquels nous avons participé, et quelques résultats des tests multimodaux.

4.3.2.1 Les mesures retenues

La première stratégie envisagée a été de fournir un ensemble de signatures rendues anonymes, et de demander une comparaison deux à deux de ces signatures. Les scores de similarité obtenus constituent la "matrice de similarité" des signatures. Cette méthode conduit à un grand nombre de comparaisons ($N^2 / 2$). Elle présente l'avantage de fournir tous les scores de toutes les comparaisons, permettant d'effectuer, *a posteriori*, des "évaluations virtuelles" sur des sous-ensembles de cette matrice ainsi que des évaluations de systèmes d'identification. Malheureusement, les partenaires ont décidé, principalement pour des raisons de complexité, de procéder plutôt par liste de comparaisons deux à deux à effectuer entre signatures. Ainsi, la génération de la liste de comparaisons est effectuée par un tirage aléatoire avec un nombre de comparaisons inter et intra-classes à peu près identiques.

Nous nous sommes donc placés dans le contexte de la vérification d'identité. Pour les comparaisons entre images; deux cas sont envisagés:

- Comparaison « intra »: entre deux images différentes d'une même personne.
- Comparaison « inter »: entre deux images de deux personnes différentes.

En se plaçant dans le cadre des systèmes de contrôle d'accès, les deux taux de références sont calculés: le taux de faux rejet (FRR : False Rejection Rate) et le taux de fausse acceptation (FAR : False Acceptance Rate). Ces deux taux sont fonction de la valeur du seuil de décision t (cf chapitre 1). Rappelons qu'un seuil t faible facilite l'accès donc conduit à FAR important et qu'un seuil t fort entraîne un FRR fort.

Ce seuil t est fixé lors de la phase de développement de manière à optimiser une fonction nommée DCF: Detection Cost Function. Le critère d'optimisation retenu peut-être une somme pondérée s'appuyant sur les deux taux FAR et FRR , les coefficients C_a et C_r permettant de privilégier l'un ou l'autre de ces taux selon l'application visée :

$$DCF(t) = C_a \cdot FAR(t) + C_r \cdot FRR(t) \quad (4.1)$$

Pour avoir un test statistiquement significatif pour obtenir la valeur de ce taux, le nombre d'erreurs constatées doit au moins être égal à une dizaine, tant pour les comparaisons « intra » que « inter ». Ainsi pour valider un EER de 5%, cela signifie qu'il faut au moins avoir pratiqué 200 comparaisons dans chacune des deux catégories. Remarquons que la DCF prise comme référence est calculée pour :

$$DCF(t) = 0,99 \cdot FAR(t) + 0,1 \cdot FRR(t) \quad (4.2)$$

De telles valeurs de C_a et C_r , induisent une application pour laquelle on souhaite avoir un taux extrêmement faible de fausses acceptations, quitte à entraîner un certain désagrément pour les clients qui se verront, dans 10% des cas, refusés à tort l'accès au site.

4.3.2.2 Evaluation visages 2D

5000 fichiers images ont été collectés à partir de différentes sources : images camescope ("dv"), images des caméras numériques constituant la tête stéréoscopique de face ("dcam"), images webcam ("wcam"). Les visages 2D sont extraits d'une manière automatique à partir des vidéos. Compte tenu du nombre important de fichiers à traiter, les fichiers visages ont été annotés en semi-automatique par l'un des partenaires (URATEK) (détection automatique de la position des yeux puis vérification visuelle, et modification si nécessaire), en se basant sur la validation de la précision des algorithmes effectuée au cours de la phase de d'évaluation préliminaire. Ceci a permis également de détecter et d'éliminer automatiquement un certain nombre de fichiers erronés (profils au lieu de visages de face, erreurs de cadrage, mouvements de mains ...). Ainsi, données brutes, annotations et données normalisées ont été fournies aux différents participants. Les données disponibles ont été utilisées pour construire plusieurs séries de tests, présentées dans le tableau 4.2. L'ensemble contenant les signatures de référence, donc celles qui sont utilisées comme base pour les comparaisons proposées, est désigné sous le nom de « Galerie ». Par ailleurs l'ensemble des signatures utilisées pour le test est désigné sous le nom de « Probe ».

N° Test	« Galerie »	« Probe »	Objectif
03	multisession s1 - dv	multisession s2 - dv	Test de base multisession
04a	session unique -dv	session unique - dv	Test de base monosession (incluant les variations d'expression)
04b	session unique -dv	session unique -dv	Test sur la variation d'éclairage (galerie: éclairage fixe, probe: éclairage variable)
05a	session unique -dv	session unique - dcam	Test sur la qualité des images (comparaison galerie dv, probe dcam)
05b	session unique -dv	session unique - wcam	Test sur la qualité d'image (comparaison galerie dv, probe wcam)

Tableau 4.2 : Séries de tests élaborées à partir des données collectées.

Ma participation à l'ensemble des tests de la modalité « Visage 2D » a été effectuée avec les deux algorithmes de référence : *Eigenfaces* (PCA) et *Fisherfaces* (LDA). Deux autres participants, dont l'un du laboratoire IBISC, et l'autre du GET-INT ont participé aussi à l'ensemble de ces tests avec les *Eigenfaces* Modulaires et les ondelettes de Gabor respectivement [MCG⁺09]. Les résultats ont été mis en comparaison avec ceux obtenus par la plateforme de référence BIOSECURE [Bio.eu]. Pour chaque test, plusieurs essais ou « runs » ont été fournis incluant des variations à propos de la distance utilisée et de la méthode de normalisation des scores.

4.3.2.3 Le déroulement des évaluations et les résultats renvoyés par l'évaluateur

L'évaluateur communique une liste des images entre lesquelles les comparaisons sont effectuées. Il donne également un ensemble d'images utilisées pour l'apprentissage des algorithmes si ceux-ci nécessitent cette phase. Les données de développement nécessaires à la mise au point des algorithmes ont quand à elles été extraites de la base de développement, qui ne correspond pas en totalité au standard d'acquisition définie par IV².

Chaque participant renvoie la mesure de similarité obtenue pour chacune des comparaisons demandées par l'évaluateur. Un participant peut renvoyer plusieurs séries de résultats pour un même test. Ces différents « runs » correspondent à l'utilisation de différentes distances, à de faibles variations entre différents algorithmes.

Une fois en possession des résultats des participants, l'évaluateur procède à une phase de normalisation de ces résultats puis renvoie à chacun d'eux des résultats quantitatifs et des analyses sous forme de courbes. Nous prenons comme exemple le test 04a incluant les variations d'expression dans les images faciales d'une seule session. Les figures suivantes 4.1 et 4.2 montrent les résultats sur les scores issus de l'algorithme *Fisherfaces* basé sur l'Analyse Discriminante Linéaire (LDA) que nous a retourné l'évaluateur.

La première série de résultats renvoyés porte sur l'analyse des scores de similarité dans les comparaisons de type « intra » et de type « inter ». Comme l'illustre l'exemple ci-dessous, pour un nombre total donné de comparaisons et un nombre de comparaisons « intra » spécifiées, on indique au participant la moyenne et les écarts-type des similarités « intra » et « inter ». Les histogrammes pour ces deux catégories de comparaisons sont également tracés (figure n°4.1).

Nombre de comparaisons : 5049
 Nombre de comparaisons « intra » : 2595
 Moyenne « intra » : 0.607 écart-type intra : 0.124
 Moyenne « inter » : 0.254 écart-type inter : 0.075

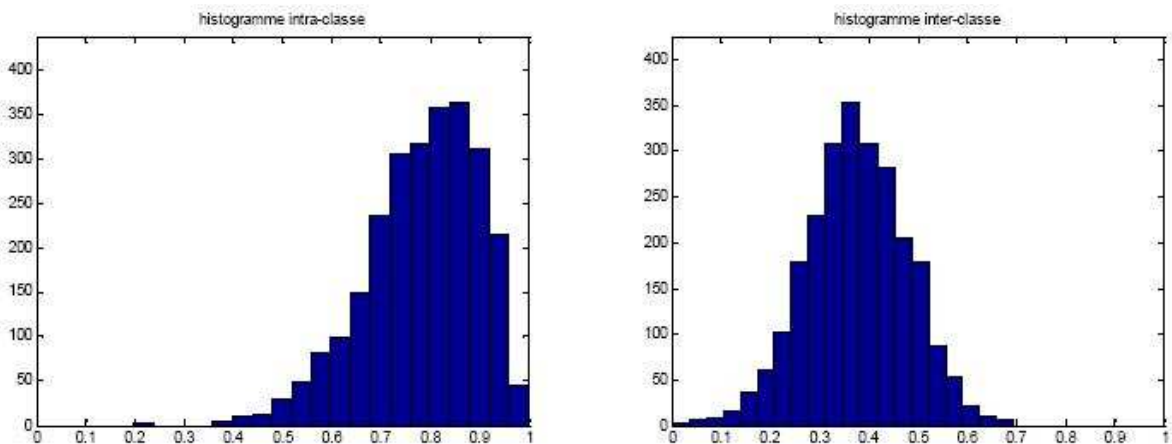


Figure 4.1 : Histogrammes des distances intra-classe & inter-classe.

La deuxième catégorie de résultats communiqués par l'évaluateur consiste en deux courbes *DET* (Detector Error Trade-off) ainsi que l'exploitation de celles-ci. Les courbes *FAR* et *FRR* tracent l'évolution des probabilités d'erreur en fonction du seuil t . Elles dépendent donc de la définition du score, et de la méthode de normalisation utilisée. En revanche, on peut tracer la courbe *DET* ($FRR(t)$) en fonction de $FAR(t)$ qui est indépendante de la définition de la mesure. Sur l'axe des abscisses on porte les taux de *FAR* et l'axe des ordonnées porte les taux de *FRR*. Ces taux sont compris entre 0 et 1. Pour une valeur du seuil t donnée, on a un couple $FAR(t); FRR(t)$, qui conduit à un point de la courbe *DET*. Ces courbes peuvent être réalisées avec une échelle linéaire ou logarithmique (cf Figure 4.2).

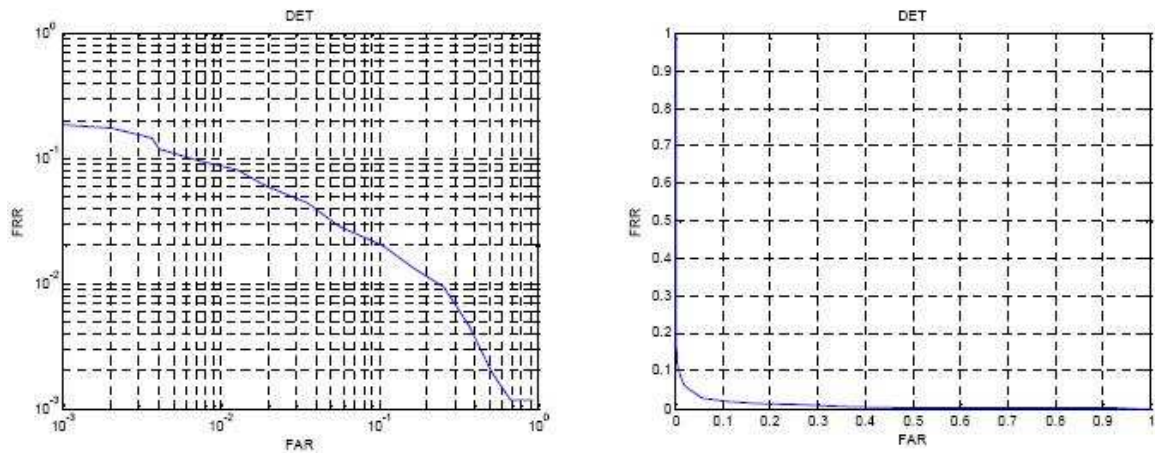


Figure 4.2 : Courbes DET réalisées avec une échelle logarithmique et linéaire.

Après normalisation des mesures de similarité communiquées par chaque participant, l'évaluateur fait varier les seuils et réalise les courbes *DET* ci-dessous. A partir de ces courbes on mesure le taux d'équivalence d'erreur (*EER*) et la valeur du taux de faux rejet (*FRR*) pour un taux de fausses acceptations (*FAR*) fixé à 1% par exemple. Ci-dessous, nous présentons les meilleurs résultats obtenus entre les différents « runs » de chaque participant et ceci au test 04a.

participant	IBISC1	IBISC2	IBISC3	INT	Biosecure
algorithme	LDA	PCA	PCA modulaire	Gabor + LDA	PCA
distance	L1	L1	L1	cosinus	cosinus
EER	0.037	0.066	0.053	0.042	0.073
FRR (FAR=0.1)	0.021	0.047	0.034	0.021	0.055

Tableau 4.3 Meilleurs scores de chaque algorithme pour le test 04a.

Ma participation à cette évaluation correspond aux algorithmes IBISC1 et IBISC2. Par ailleurs, afin de permettre la comparaison des résultats obtenus par les différents participants sur un même test, une courbe *DET* synthétique est présentée sur la figure 4.3.

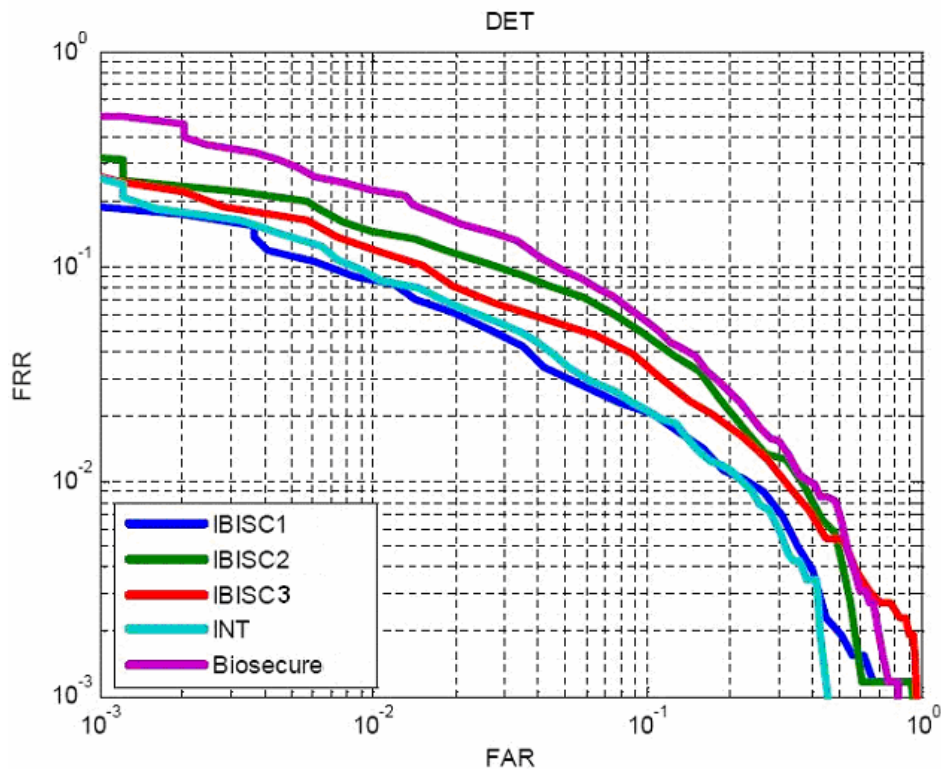


Figure 4.3 : Exemple d'une courbe DET regroupant les différents résultats des participants à un même test (ici visages 2D et Test03).

4.3.2.4 Evaluation multimodale

La base de données IV² est multimodale. En complément des tests d'algorithmes multimodaux effectués sur la base des visages parlants, des algorithmes de fusion de scores ont été appliqués aux résultats des trois modalités de visage 2D, visage 3D et iris. La fusion de scores se décompose en deux principales étapes : La normalisation des données et la fusion elle-même. Une phase préliminaire a été menée pour permettre de définir un protocole pour ces évaluations multimodales. Suite à cela, il a été décidé d'utiliser les résultats des algorithmes des différents participants sur les listes de tests communes aux diverses modalités évaluées ici. Les fichiers de test fournis aux participants sont constitués de listes de comparaisons accompagnées pour chacune de trois scores (visage 3D, visage 2D, iris).

Comme dans les évaluations monomodales, deux séries de tests ont été proposées aux participants :

- Test03 est un test multi-sessions, pour lequel l'ensemble « Galerie » a été pris sur les enregistrements de la première session et l'échantillon à tester (ensemble « Probe ») sur les enregistrements des secondes ou troisièmes sessions. La base d'évaluation du Test03 est constituée de 1470 comparaisons. La base d'apprentissage est constituée de 3690 comparaisons.
- Test04 est un test mono-session, plus facile a priori, mais comportant un échantillon de taille plus élevée. Afin de ne pas biaiser les résultats, les identités utilisées pour le test 04 ont été séparées en deux catégories : environ 1/3 des personnes (soit 88 identités) ont servi à l'apprentissage des méthodes de fusion alors que les 2/3 restants (soit 176 identités) ont servi à leur évaluation. Au final, l'ensemble d'apprentissage contient donc 712 comparaisons et l'ensemble d'évaluation 1831.

Pour cette évaluation des performances des systèmes biométriques multimodaux, deux participants ont testé leurs approches, le GET-INT avec 3 algorithmes et Thales avec 4 algorithmes.

4.3.3 Conclusion sur les évaluations

Pour conclure sur les évaluations réalisées, le tableau ci-dessous donne une idée de l'importance du travail fourni. Même si on peut déplorer le faible nombre de participants par modalité, on peut constater le grand nombre de comparaisons soumises par test et le nombre important de « runs » différents rendus par chaque participant.

Le rapport d'évaluation [CH07] propose un certain nombre d'analyse des résultats obtenus par chacun des algorithmes à travers une même modalité, ainsi qu'un retour plus global sur les différentes modalités.

Ce projet nous a permis de valider le bon fonctionnement des algorithmes *Eigenfaces* (PCA) et *Fisherfaces* (LDA) qui seront utilisés dans les chapitres 5 et 6 du présent document.

<i>Modalités</i>	<i>Référence du test</i>	<i>Nombre de comparaisons soumises</i>	<i>Nombre de participants</i>	<i>Nombre de runs</i>	<i>Total des comparaisons pour la modalité</i>
Visages 2D	Test 03	3592	4	16	318272
	Test04a	5049	4	16	
	Test04b	4867	4	16	
	Test05a	3132	4	16	
	Test05b	3252	4	16	
Multimodalité	Test 03a	1470	2	16	105632
	Test 03b	1470	2	16	
	Test04a	1831	2	16	
	Test04b	1831	2	16	

Tableau 4.4 Tableau récapitulatif des différents tests pour les évaluations du visage 2D et de la multimodalité réalisées.

4.4 Les bases de visages utilisées

Plusieurs bases de visages ont été développées pour l'évaluation des algorithmes de reconnaissance faciale. Chacune comporte des conditions de prises de vues différentes. Les bases les plus anciennes (ORL et YALE) ont été le plus utilisées et permettent de comparer plus facilement de nouvelles méthodes à celles de l'état de l'art. Les plus récentes (Color FERET, FRGC, CVL, AR et IV²) contiennent plus de personnes et sont donc utiles pour des évaluations à plus grande échelle. D'autres bases de visages sont disponibles et destinées à des évaluations adaptées à certaines variabilités du visage telles que les bases UMIST, BANCA, PF01, Yale et PIE. Ces trois dernières bases par exemple (PF01, Yale et PIE) disposent d'un nombre important de poses différentes mais renferment seulement quelques dizaines de personnes acquises lors d'une seule session.

Différents facteurs sont appliqués sur les visages à savoir des changements d'éclairage, de poses, d'expressions faciales et des occultations. Les variations dans le temps des visages sont étudiées à travers l'acquisition de plusieurs sessions avec un intervalle de durée défini. Nous nous sommes intéressés spécialement aux bases de données contenant plusieurs images de face par personne pour développer notre corpus de travail. Nous avons toléré des rotations de $\pm 20^\circ$ ce qui n'affecte pas sensiblement les performances des algorithmes de reconnaissance d'après l'état de l'art. Notre choix s'est arrêté sur les bases XM2VTS [Mes99], Color FERET, AR, CVL et ORL en plus de la base IV² pour développer notre corpus de travail. Nous détaillons principalement ces bases d'images ci-dessous. Une liste plus complète et plus détaillée est disponible dans [Gro04].

4.4.1 La base XM2VTS [Mes99]

La base XM2VTS (*The Extended M2VTS Database*) est une base de données multimodale qui renferme des images faciales 2D et 3D, des séquences vidéo avec des enregistrements de la voix. La base de données a été acquise dans le cadre du projet M2VTS (*Multi Modal Verification for Teleservices and Security applications*) à l'Université de Surrey (UK). La base de données contient 295 personnes enregistrées en 4 sessions durant une période de 4 mois. La base de données de visage 2D contient des prises de vue faciales et de profil. Deux prises de vue faciales ont été enregistrées, à chaque session, ce qui fait un total de 8 images pour chaque personne. La figure 4.4 illustre un exemple d'images d'une personne. Les images en colonnes représentent les sessions auxquelles elles appartiennent.



Figure 4.4: Exemples d'images de la base XM2VTS.

La base de données XM2VTS a été conçue initialement pour l'authentification de personnes. Ainsi un protocole d'évaluation a été élaboré spécifiant les ensembles d'apprentissage, d'évaluation et de test [LM98]. Nous revenons sur ce protocole dans le chapitre suivant pour évaluer notre système d'identification.

4.4.2 La base FERET

La base de données FERET a été collectée dans le cadre du programme « Facial Recognition Technology » [PMRR00, Nist.gov] mené par le National *Institute of Standards and Technology* (NIST) Américain. Il s'agit de la plus grande base disponible pour les chercheurs qui a été acquise avec des poses différentes et durant 15 sessions entre 1993 et 1996. Les images, initialement collectées depuis un appareil photographique de 35mm ont ensuite été digitalisées. Une première version de cette base de données a été réalisée en 2001 et contient 14051 images faciales en niveaux de gris avec une résolution de 256 x 384 pixels. La version la plus récente, réalisée en 2003, renferme des images numériques couleurs de plus grande qualité avec une résolution de 512 x 768 pixels et une compression sans perte de données à la différence des premières images en niveaux de gris. En plus, de multiples erreurs de noms d'images, d'identifiants, et dates de capture, qui figurent sur la première base à niveau de gris ont été corrigées. Cette dernière base contient 11338 images représentant 994 personnes

différentes. Pour chaque individu, on dispose d'une vue faciale régulière *fa* et une vue faciale alternative *fb* prise un peu après *fa*. D'autres poses ont été acquises pour la majorité de ces individus allant du profil gauche au profil droit avec des rotations de 15°, 22°, 45°, 67°, et 90° en profondeur de la tête. Pour quelques personnes de la base, on dispose d'autres vues « *duplicate* » collectées dans des conditions similaires à *fa* et *fb* mais dans des sessions ultérieures. 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 540 jours après la première prise de vue.

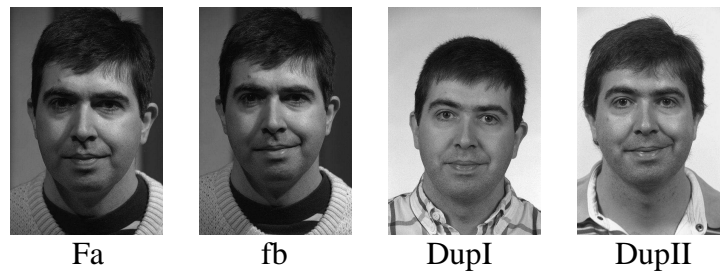


Figure 4.5 : Extrait de la base Color FERET
Les images sont transformées en niveau de gris.

4.4.3 La base AR [MB98]

La base AR 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 x 576 pixels. 13 vues de chaque sujet ont été collectées. Pour la majorité de ces personnes, 13 autres vues ont été acquises lors d'une deuxième session à deux semaines d'intervalle. Un extrait des images collectées lors de la première session est donné en figure 4.6. Ces vues renferment des changements d'expression faciale, d'éclairage, ainsi que des occultations partielles des yeux (lunettes de soleil) et de la partie basse du visage (cache-col). Lors de la seconde session, les 13 vues sont collectées dans les mêmes conditions que pour la première.

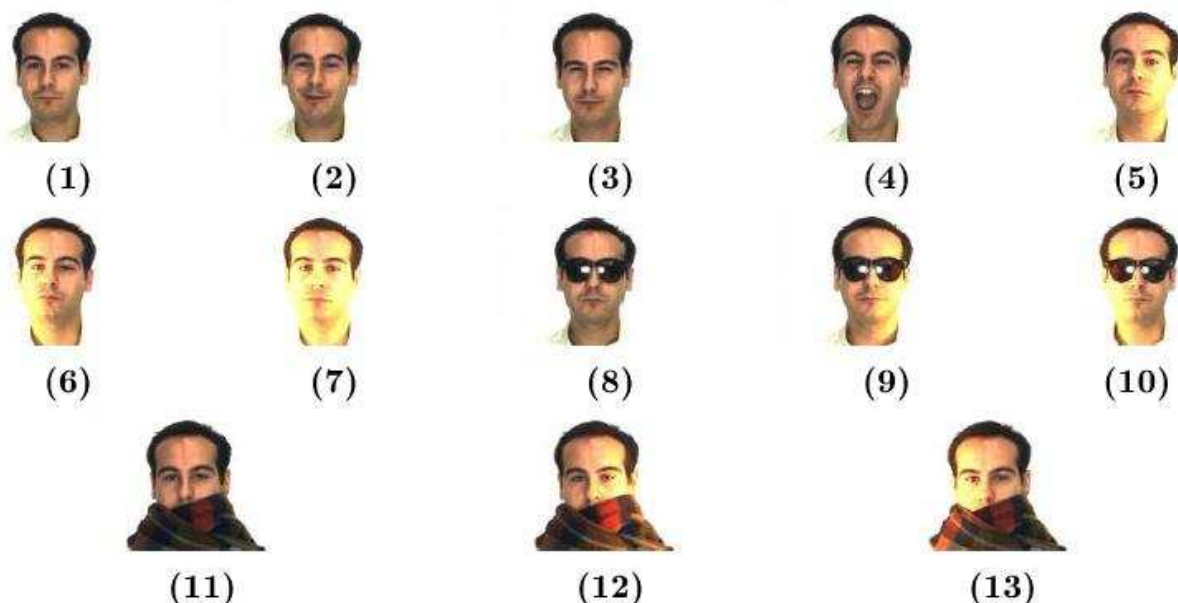


Figure 4.6 : Extrait de la base AR. Ensembles des vues collectées pour un individu de la base.

4.4.4 La base CVL [CVL.si]

Cette base de données de visage a été créée au *Computer Vision Laboratory (CVL)* à l'université de Ljubljana [CVL.si]. 114 personnes y sont enregistrées. Les images sont en couleur de taille 640 x 480 pixels. 7 vues de chaque sujet ont été collectées. Ces vues renferment différentes poses et expressions prises dans des conditions de luminosité uniformes. Parmi ces images, trois seulement sont de face. La figure 4.7 illustre un exemple d'images de face de la base CVL.



Figure 4.7 : Images faciales extraites de la base CVL.

4.4.5 La base ORL

La base ORL a été collectée entre avril 1992 et avril 1994 par un laboratoire de AT&T, basé à Cambridge. La base contient 40 personnes, chacune étant enregistrée sous 10 vues différentes (cf. figure 4.8). Les images sont de taille 112 x 92 pixels. Pour quelques sujets, les images ont été collectées à des dates différentes, avec des variations dans les conditions d'éclairage, les expressions faciales (expression neutre, sourire et yeux fermés) et des occultations partielles par les lunettes. 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. Cependant, ces variations ne portent que sur certaines personnes et ne sont donc pas systématiques.



Figure 4.8 : 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'éclairage.

4.4.6 Notre corpus d'évaluation

Dans un premier temps, n'ayant pas encore la totalité de ces bases, nous avons travaillé exclusivement sur la base XM2VTS. Cette première phase d'évaluation nous a permis d'étendre rapidement nos algorithmes sur le corpus constitué par un mélange de la plupart des bases présentées. Le tableau 4.6 récapitule les différentes bases utilisées et le nombre de sessions et d'images disponibles pour chaque personne.

Concernant la base IV², nous avons récupéré les images « cam » du camescope utilisé lors de la campagne d'évaluation. 9 images par personne sont extraites lors de la première session et seulement cinq pour la deuxième session.

Base	IV ²	XM2VTS	Color FERET	AR	CVL	ORL
# de personnes	302	295	994	136	113	40
# de sessions	1-3	4	1-15	1-2	1	1
# d'images faciales par personne	9-14	8	2-22	13-26	3	10

Tableau 4.5 : Bases de données de visage formant le corpus de notre évaluation.

Ainsi nous avons regroupé une base de travail assez conséquente de 1880 personnes pour évaluer nos approches d'identification. Pour la majorité des images la position des centres des yeux était fournie avec la base. Nous avons annoté manuellement le reste des images pour pouvoir les normaliser. Ainsi, toutes les images de notre corpus sont normalisées selon la procédure décrite à la section 2.7. Notre procédure consistait à une rotation du visage dans l'image de manière à ce que l'axe interoculaire soit horizontal, suivi d'un redimensionnement des visages pour avoir une même distance interoculaire. Ces deux premières étapes qui nécessitent les coordonnées des centres des yeux dans l'image brute sont suivies d'un découpage et d'une égalisation de l'histogramme du visage découpé.

4.5 Conclusion

Le projet IV² nous a permis de travailler sur plusieurs aspects techniques pour maintenir et paramétrer les différents capteurs pour l'acquisition des différentes données biométriques ainsi que pour contrôler l'éclairage d'une manière automatique suivant un protocole bien défini. Au terme des campagnes d'acquisition, un travail très important a été effectué par nous-même et d'autres partenaires du projet pour vérifier, corriger et organiser les données acquises afin de préparer une version exploitable de la base. D'autre part, ma participation à la première campagne d'évaluation sur les données IV² m'a permis de comprendre plusieurs problématiques et défis liés non seulement à la reconnaissance des visages 2D mais aussi à la reconnaissance des visages 3D, de l'iris et de la voix. Bien que l'évaluation des systèmes d'identification n'a pas été retenue par l'ensemble des partenaires, ce projet nous a permis de valider le bon fonctionnement des algorithmes *Eigenfaces (PCA)* et *Fisherfaces (LDA)* qui seront utilisés dans les chapitres 5 et 6. Cette décision des partenaires du projet témoigne encore une fois de la complexité de la tâche de l'identification qui consiste à comparer l'ensemble des images de test avec l'ensemble des images de la galerie. En conclusion, nous pouvons dire que le projet IV² a atteint ses objectifs qui étaient de réaliser le matériel permettant d'acquérir une base de données biométriques multimodales respectant un protocole précis et de concevoir et mettre en œuvre des campagnes d'évaluation d'algorithmes de reconnaissance biométriques dans des conditions variées.

Les 300 identités de la base IV², ajoutées aux données provenant d'autres bases Benchmark (XM2VTS, FERET, AR, ORL et CVL) nous ont permis de constituer un large corpus de travail de 1880 individus. Ce corpus nous permettra d'évaluer plus rigoureusement les performances de notre approche d'identification que nous détaillons dans le chapitre suivant.

Chapitre V

Nouvelle approche d'identification dans les bases de données biométriques - Application à la reconnaissance faciale 2D

5.1 Introduction

Nous proposons dans ce chapitre une nouvelle approche d'identification dans les bases de données biométriques. Le processus classique d'identification consiste à comparer la signature de l'identité recherchée avec toutes celles de la galerie. Cette tâche se trouve complexe et très coûteuse notamment dans les grandes bases de données biométriques. Le système que nous proposons a pour objectif de simplifier la complexité et d'améliorer les performances de l'identification biométrique. Pour cela, nous avons ajouté au schéma classique d'identification une phase de partitionnement de la galerie en amont de la phase de recherche. Nous proposons une partition de la base de données suivant la similarité et la dissimilarité des caractéristiques les plus pertinentes que l'on puisse extraire de la modalité biométrique utilisée. De ce fait, notre système est fondé sur deux questions essentielles :

- l'extraction des signatures qui soient les plus discriminantes entre les individus tout en gardant leur stabilité pour les différents échantillons d'une même personne ;
- la classification de ces signatures pour regrouper les personnes les plus similaires dans les mêmes classes et séparer celles qui sont différentes.

Nous étudions dans ce chapitre plusieurs voies de recherches relatives à l'apprentissage des données pour l'extraction de signatures, à la taille des vecteurs caractéristiques extraits, au nombre de classes calculé, ... pour évaluer correctement l'approche proposée. Nous avons développé essentiellement, l'identification à partir des algorithmes *Eigenfaces*, *Fisherfaces* et moments de Zernike sur les images de visages 2D. Dans ce chapitre, nous utilisons la base XM2VTS pour évaluer les différentes capacités de notre système.

5.2 Les différentes étapes de notre approche d'identification

L'architecture générale de notre approche avec les différents processus proposés pour élaborer l'identification biométrique des individus est illustrée dans la figure 5.1. Nous avons développé principalement une étape de partitionnement pour diviser la base de données en plusieurs ensembles séparés. Cette étape de classification a pour objectif de regrouper les données (les identités) qui ont des caractéristiques similaires et à séparer celles qui ont des caractéristiques dissemblables dans des classes différentes. Nous proposons, à travers ce partitionnement des identités de la galerie, de simplifier le module d'identification. L'élargissement des partitions les plus dissemblables et ne présentant pas de caractères communs avec celles de l'identité requête permet de simplifier la recherche dans la galerie. Nous voulons entendre par

cette simplification la diminution du nombre d'identité de la galerie sur lesquelles on effectue l'étape finale de l'identification par mesure de similarité.

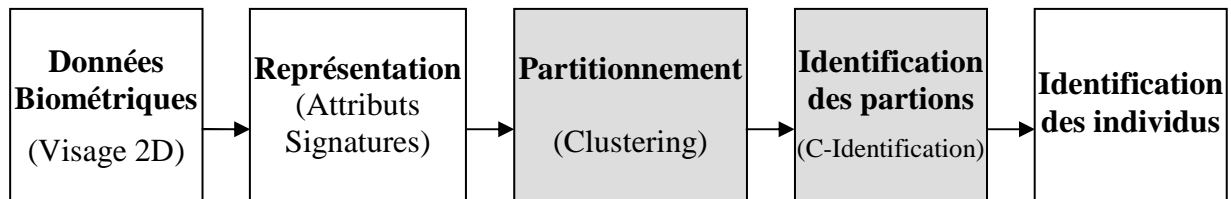


Figure 5.1 : les différentes étapes de notre approche d'identification biométrique. Les étapes en gris constituent les originalités proposées par rapport aux systèmes d'identification classiques.

Le partitionnement de la galerie étant élaboré sur des critères bien définis, nous cherchons à identifier tout d'abord la partition (constituée d'une ou de plusieurs classes) d'identités potentielles à laquelle la personne à reconnaître est supposée appartenir. Nous dénotons cette étape d'identification des classes par C-Identification pour la différencier de l'identification des individus dans un ensemble de données. La C-Identification consiste donc à déterminer les distances de la signature extraite du visage requête avec le centre ou bien l'élément représentatif de chaque classe d'individus. Les classes les plus proches constituent la partition retenue pour la suite de la recherche. L'identification finale consiste à chercher les identités les plus proches de la personne requête parmi le sous ensemble de la galerie correspondant à la partition d'identités retenues. Ceci va permettre de simplifier considérablement la tâche d'identification dans la base de données biométrique.

Plusieurs travaux, dans ce contexte de simplification des bases de données, ont été réalisés pour l'identification des individus par leurs empreintes digitales. Comme nous l'avons mentionné dans le chapitre 1, les empreintes digitales présentent des caractéristiques morphologiques globales qui aident à les discriminer en cinq classes bien définies. La majorité des systèmes aboutis de reconnaissance d'empreintes digitales emploie des techniques de classification supervisée pour séparer les empreintes sur les cinq classes prédéfinies. L'identification d'une empreinte requête est effectuée, ensuite, dans un sous ensemble de la galerie qui correspond à sa classe d'appartenance.

Nous développons une approche de recherche similaire à ces systèmes de reconnaissance d'empreintes digitales. Mais, à la différence de ces méthodes, nous proposons une stratégie de classification et de regroupement des caractéristiques biométriques représentatives des identités par leurs similarités à travers les méthodes de classification non supervisée (dites aussi de classification automatique). Nous regroupons et séparons les identités dans les classes suivant la similarité des signatures extraites. De ce fait, étant donnée une modalité biométrique, l'extraction d'attributs pertinents qui discriminent les identités en vue de leur classification et qui les indexent efficacement en vue de leur identification est une tâche critique.

Après l'extraction des signatures pertinentes (appelées aussi vecteurs caractéristiques) à partir des modalités biométriques retenues, vient ensuite la classification automatique de ces vecteurs caractéristiques. Nous explorons, tout d'abord, les structures classificatoires pour viser les méthodes qui fournissent un résultat de partitionnement de classes. Ces différentes méthodes de classification automatique telles que les K-moyennes et les centres mobiles de Forgy possèdent des modèles d'apprentissage spécifiques de ces structures de partitionnement. D'autre part, différentes questions relatives au nombre de classes et à la qualité de la

classification sont difficiles à relever surtout en présence de données implicites représentant une modalité biométrique donnée. Nous développons, néanmoins, dans ce travail plusieurs configurations et stratégies classificatoire pour optimiser le partitionnement des attributs biométriques et faciliter l'indexation des identités relatives.

5.3 Les données biométriques

Les données biométriques à l'entrée d'un système d'identification doivent être acquises sans contraintes imposées aux utilisateurs et aux personnes à identifier. Alors que les systèmes d'authentification conçus essentiellement pour des applications de contrôle d'accès peuvent utiliser une large panoplie de modalités biométriques, les systèmes d'identification biométrique visent principalement des applications de surveillance et d'investigation criminelle et nécessitent donc des modalités biométriques faciles à enregistrer et avec une coopération réduite de la personne. La reconnaissance de visages est l'une des techniques privilégiés dans ce contexte, parce qu'elle est non intrusive et à cause de la relative facilité d'acquisition de photographies, pouvant même être effectuée à l'insu de la personne concernée. Cette facilité d'acquisition des visages rend la constitution d'une base de données assez aisée pour les applications d'identification.

5.3.1 Les métadonnées

Les métadonnées sont des données qui servent à caractériser les données biométriques acquises. Les métadonnées peuvent être des renseignements personnels sur l'utilisateur tels que la couleur des yeux, l'âge, la taille etc., ou des informations de contexte liées aux conditions d'acquisition à un moment précis (température, luminosité, bruit etc.). Les métadonnées sont donc des données de nature différente de celles de l'image ou du signal acquis, qui doivent soit être renseignées par l'utilisateur ou le contrôleur, soit être acquises automatiquement par un dispositif d'acquisition complémentaire du système d'acquisition des données biométriques.

Ces métadonnées peuvent permettre d'adapter le système à un type de personne (tranche d'âge, couleur de la peau, ...) ou à la précision des conditions d'acquisition pour spécifier des paramètres de fusion et de décision par exemple [RNJ06, All09]. Pour développer notre système d'identification, nous n'avons pas inclus de tels processus. Cependant, nous avons utilisé uniquement les centres des yeux comme métadonnées pour élaborer l'étape de normalisation des images faciales. Généralement la normalisation consiste en une rotation du visage dans l'image de manière à ce que l'axe interoculaire soit horizontal, suivi d'un redimensionnement des visages pour avoir une même distance interoculaire. Les deux premières étapes du processus de normalisation se basent sur la position des centres des yeux dans l'image brute.

5.3.2 Organisation des données

Pour réaliser un système d'identification fiable et robuste aux aléas des conditions réelles des applications, les images d'évaluation doivent être organisées en trois ensembles séparés : la base d'apprentissage, la base d'enrôlement et la base de test comme l'illustre la figure 5.2.

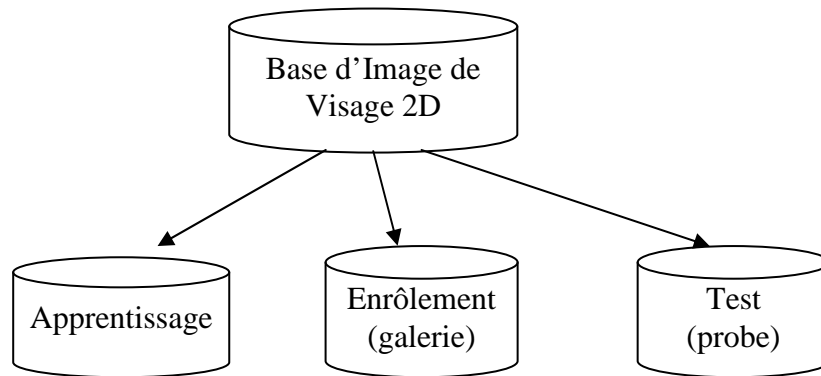


Figure 5.2 : Organisation des données pour la reconnaissance faciale 2D.

La base d'apprentissage est conçue pour construire un nouvel espace réduit des visages sur lequel sont représentées (projetées) les images faciales. La représentativité et la robustesse du nouvel espace de visages dépendent essentiellement de la taille (du nombre d'images) de cette base d'apprentissage ainsi que de la variété des images en termes de facteurs pouvant altérer l'apparence globale du visage comme les expressions et/ou la luminosité.

Les images d'enrôlement correspondent aux individus enregistrés réellement dans la base de données. Elles sont utilisées pour créer les vecteurs caractéristiques et les modèles (*template*) qui vont servir comme un code d'identification. Généralement un modèle, d'une personne, est créé à partir de l'ensemble de ses signatures déduites des différentes images disponibles dans la base d'enrôlement. La robustesse et la qualité d'une méthode de représentation (PCA, LDA, ...) se mesure à travers la compacité des signatures et leur facteur discriminant : deux images semblables génèrent deux signatures assez proches au sens d'une mesure de similarité et vice versa. Pour entraîner le module de classification non supervisé des signatures biométriques extraites, nous utilisons également la base d'enrôlement pour générer les classes et les partitions.

Les images de test sont projetées à leur tour sur l'espace de représentation des visages. Les signatures extraites de ces images de test sont appariées avec les signatures enregistrées dans la base de données extraites lors de la phase d'enrôlement. Les résultats d'appariement fournissent finalement une décision d'identification et/ou de reconnaissance.

5.3.3 Stratégies expérimentales

Nous avons suivi, initialement, le protocole de Lausanne relatif à la base de données XM2VTS [LM98, Mes99] pour évaluer les performances de notre approche. Nous avons utilisé l'ensemble de clients, déjà établi, constitué de 200 personnes et nous avons opéré l'identification dans un monde fermé. Chaque personne a 8 images faciales : 4 sessions différentes avec 2 images pour chaque session (voir figure 4.16). Nous avons divisé cet ensemble en trois sous-ensembles d'apprentissage, d'enrôlement et de test. Les images d'apprentissage sont utilisées pour construire l'espace de visage notamment par la méthode des *Eigenfaces* et celle des *Fisherfaces*. La troisième méthode de représentation des visages par les moments de Zernike que nous avons développés calcule directement des attributs géométriques à partir d'une image et ne nécessite pas la constitution d'une base d'apprentissage. Ensuite, les images d'enrôlement sont sélectionnées pour extraire les vecteurs caractéristiques des individus supposés être dans la galerie qui constitue la base de

données du système. Finalement, les images de tests sont sélectionnées pour simuler le processus d'identification.

Nous avons conçu quatre subdivisions différentes, appelées experts, pour mesurer l'impact de la phase d'apprentissage sur les performances d'identification (voir tableau 5.1). Les experts diffèrent uniquement par leurs ensembles d'apprentissage. Chaque expert apprend différemment l'espace de visage sur lequel seront projetées les images d'enrôlement pour extraire leurs vecteurs caractéristiques. D'autre part, nous avons conçu le même ensemble d'enrôlement constitué de trois images par personne et le même ensemble de test avec deux images par individu. Les références exactes des images sélectionnées pour ces deux ensembles d'enrôlement et de test sont explicitées dans le tableau 5.2 qui détaille les partitions du premier expert. Ainsi, nous avons une base d'enrôlement et une base de test communes aux différents experts contenant respectivement 600 et 400 images.

Expert 1	Expert 2
Apprent. - Enrôl. - Test 600 - 600 - 400	Apprent. - Enrôl. - Test 800 - 600 - 400
Expert 3	Expert 4
Apprent. - Enrôl. - Test 1200 - 600 - 400	Apprent. - Enrôl. - Test 1600 - 600 - 400

Tableau 5.1 : Tailles respectives des bases d'apprentissage, d'enrôlement et de test pour chacun des 4 experts.

Nous donnons ici les détails du premier expert qui assure une séparation complète entre les trois ensembles. La première image de chacune des trois premières sessions est utilisée pour l'apprentissage et la construction de l'espace de représentation des visages (voir Tableau 5.2). La deuxième image de chacune de ces trois premières sessions est utilisée pour enrôler les identités dans la galerie. Enfin, les deux dernières images de la quatrième session sont utilisées pour les tests de classification et d'identification. Ainsi, nous disposons d'un ensemble de 600 images d'apprentissage, un ensemble de 600 images d'enrôlement et un ensemble de 400 images de test comme l'illustre le tableau 5.1.

Expert 1		Session			
		<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
image	<i>1</i>	Apprentissage	Apprentissage	Apprentissage	Test
	<i>2</i>	Enrôlement	Enrôlement	Enrôlement	Test

Tableau 5.2 : Stratégie du choix des images pour les trois bases du premier expert.

Notons pour finir que pour les trois premiers experts les images des ensembles d'apprentissage sont extraites à partir des trois premières sessions. Pour sa part, la base d'apprentissage du quatrième expert comprend l'ensemble des images des quatre sessions.

5.3.4 Résultats d'identification classique

Nous avons effectué l'identification classique avec les *Eigenfaces* pour ces quatre dernières stratégies d'apprentissage de l'espace de visage. Nous représentons sur la figure 5.3 les courbes de scores cumulés (taux d'identification cumulés) pour ces quatre experts.

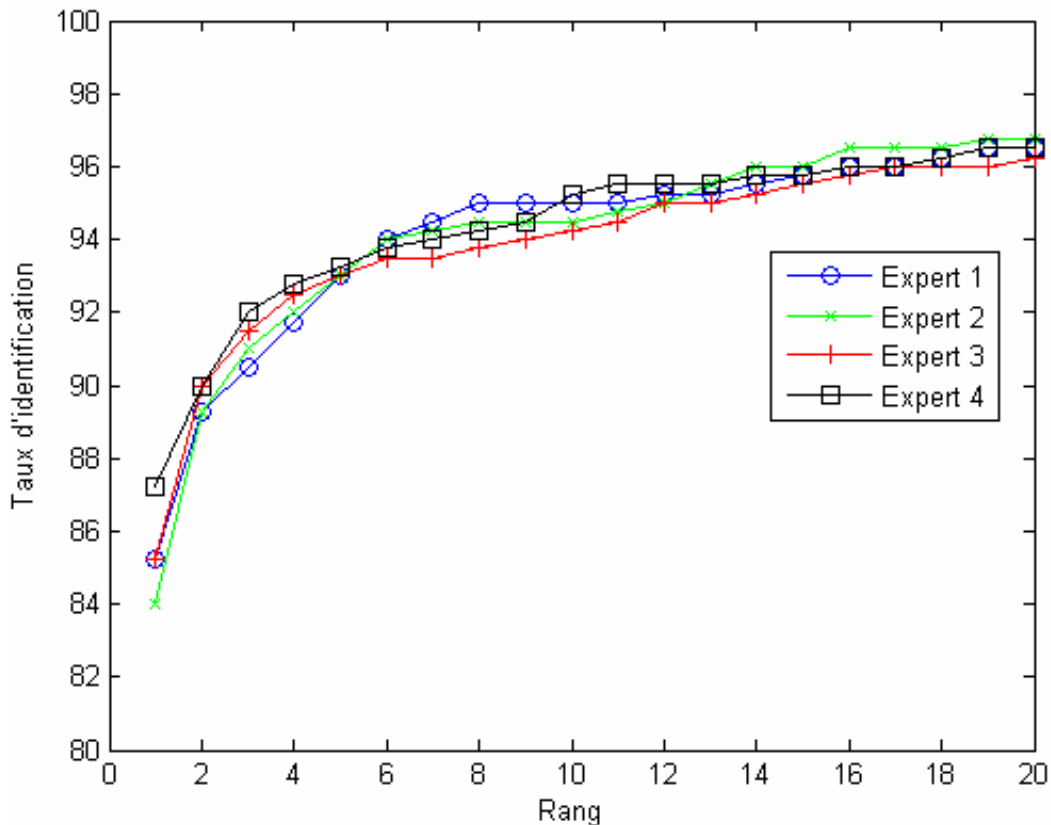


Figure 5.3 : Courbes de Scores Cumulés des *Eigenfaces* entraînées différemment avec des ensembles de 600, 800, 1200 et 1600 images de la base XM2VTS.

Nous remarquons d'après ces courbes de scores cumulés que les taux d'identification sont quasiment dans la même fourchette de 84 à 85%. Le quatrième expert donne de meilleurs résultats d'identification du fait de la richesse de sa base d'apprentissage. Cependant, puisque la base d'apprentissage de ce dernier expert englobe toutes les images d'enrôlement et de test, son taux d'identification égal à 87,25% est forcément biaisé. D'autre part, en visualisant le taux d'identification à un rang donné, tous les experts sont équivalents et aucune différence significative de résultats n'est observée.

Un autre point important pour évaluer un système de reconnaissance faciale est le nombre d'images par personne qui sont utilisées lors de l'enrôlement des individus dans la galerie. Une moyenne simple des signatures extraites de chacune des images d'un individu est calculée pour générer son modèle représentatif. Selon Tana et Songcan [TS06], plus le nombre d'images d'enrôlement par individu est élevé, meilleures sont les performances du système de reconnaissance. Pour évaluer ce point séparément, nous avons repris le deuxième, le troisième et le quatrième experts en leur imposant la même base d'apprentissage que le premier expert. Par contre une base d'enrôlement différente est sélectionnée pour chacun d'eux. Nous avons enrôlé respectivement 3, 4, 6 et 8 images par personnes pour les différents experts respectivement. Le quatrième expert se trouve encore biaisé du fait que la base

d'entraînement est confondue en partie avec la base d'apprentissage et surtout avec la base de test. La signature extraite pour chaque individu de la galerie est une moyenne des attributs extraits à partir de toutes ses images d'entraînement. Le tableau suivant explicite le nombre d'images utilisé pour chaque configuration.

Expert 1'	Expert 2'
Apprent. - Enrôl. - Test 600 - 600 - 400	Apprent. - Enrôl. - Test 600 - 800 - 400
Expert 3'	Expert 4'
Apprent. - Enrôl. - Test 600 - 1200 - 400	Apprent. - Enrôl. - Test 600 - 1600 - 400

Tableau 5.3 : Variation de la taille de la base d'entraînement pour chacun des quatre experts.

L'effet de l'augmentation du nombre d'images lors de la phase d'entraînement est bien visible contrairement à l'expérimentation précédente relative à la variation de l'ensemble d'apprentissage. Les résultats illustrés dans [TS06] sont bien validés par cette expérimentation. Plus on utilise d'images pour extraire un modèle d'une personne donnée, plus ce modèle est stable et représentatif de cette personne compte tenu de la variabilité de l'expression, de l'éclairage et des poses.

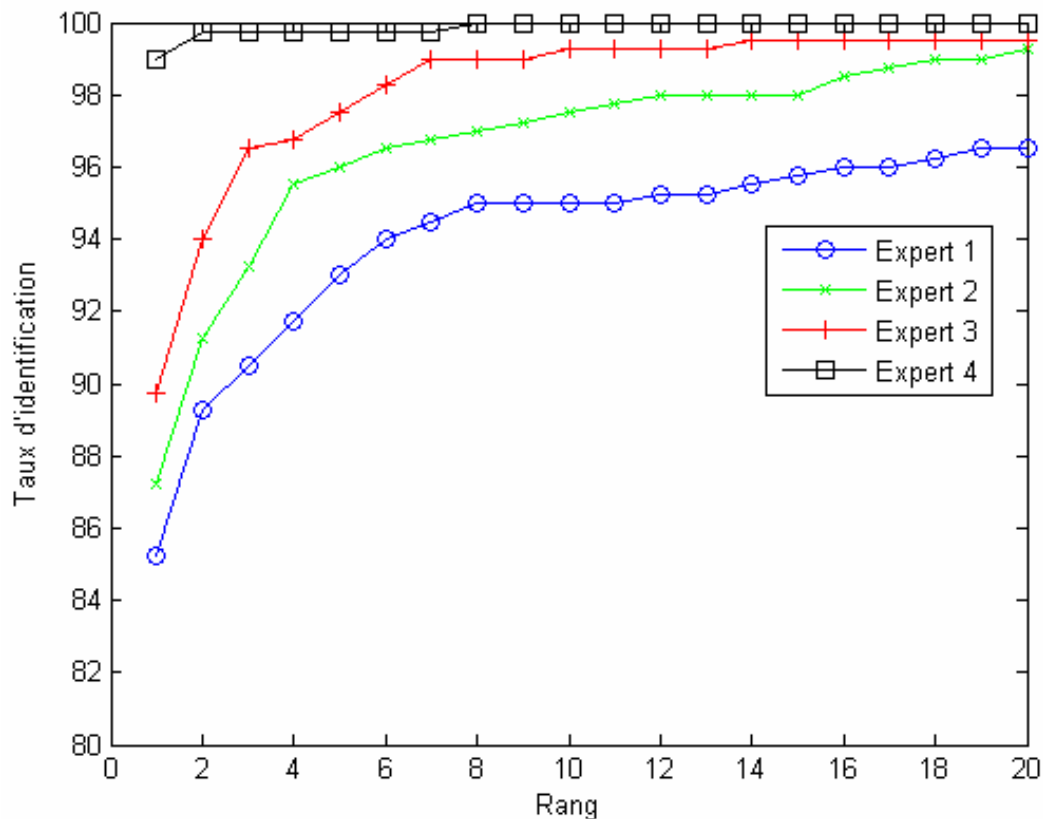


Figure 5.4 : Courbes de Scores Cumulés des Eigenfaces avec différents ensembles d'entraînement.

5.4 Représentation des visages 2D

Dans cette section nous développons une boîte à outils de différentes méthodes représentant chacune différemment l'information visuelle du visage 2D. L'utilisation d'algorithmes différents (multi-algorithmique) représentant aussi indépendamment que possible les informations présentes dans les images de visages 2D est nécessaire entre les deux processus de classification et d'identification. Notre objectif étant de générer plusieurs signatures différentes de chaque personne de la galerie pour pouvoir partitionner les identités suivant un type de signature différent de celui de l'identification. L'indépendance entre ces signatures conditionnera éventuellement les résultats finals d'identification après le processus proposé de simplification de la base de données par classification. Nous développons dans cette section les différentes méthodes de représentation retenues, à savoir les méthodes factorielles des *Eigenfaces* et des *Fisherfaces* et la méthode des moments de Zernike. Nous cherchons à extraire les traits pertinents pour les deux processus de classification automatique et d'identification.

5.4.1 Les méthodes factorielles

Soit l'espace I , l'espace des images, contenant l'ensemble des vecteurs images de visages. Les dimensions de cet espace, qui correspondent à la taille des images, sont très importantes, ce qui rend la classification difficile. Les images de visages partagent un certain nombre de propriétés structurelles communes. Les visages ne sont pas distribués de manière aléatoire dans I et une grande partie de cet espace ne correspond pas à des visages. De plus, dans le cas des visages, l'information contenue dans I est souvent redondante. Par conséquent, on peut considérer que les visages appartiennent à un sous espace F de I , de dimension inférieure, appelé espace des visages [SM04]. Les méthodes factorielles, aussi appelées méthodes des sous-espaces, visent à définir ce sous espace pour mettre en correspondance les visages à l'intérieur de ce sous espace.

Les méthodes linéaires visent à définir la nouvelle base de l'espace des visages et à éliminer les vecteurs de base les moins porteurs d'information au sens d'un critère bien choisi [MDD06]. Ainsi, nous pouvons construire la base de l'espace des visages de dimension réduite F . La dimension g intrinsèque du sous-espace F est donc fixée *a posteriori*. Les principales méthodes linéaires utilisées dans le contexte de la reconnaissance de visages sont : l'Analyse en Composantes Principales (*PCA*) notamment utilisée par la méthode des *Eigenfaces*, et l'Analyse Discriminante Linéaire (*LDA*) utilisée par la méthode des *Fisherfaces*.

5.4.1.1 Les *Eigenfaces*

L'analyse en Composantes Principales définit le sous espace F de l'espace initial I tel que la dispersion des données, projetées orthogonalement dans F , soit maximale.

Soit $\{A_1, A_2, \dots, A_N\}$ l'ensemble des vecteurs images de la base d'apprentissage. Ces données sont centrées tel que l'on a :

$$\bar{A} = \frac{1}{N} \sum_{l=1}^N A_l = 0 \quad (5.1)$$

La matrice de covariance C est ensuite calculée suivant : $C = AA^T$. La matrice Φ constituant la base orthonormée de F est résolue par l'équation 2.11.

Les colonnes de Φ sont constituées des g vecteurs propres orthonormés de la matrice de covariance des données C associés aux plus grandes valeurs propres. La valeur propre associée à chaque vecteur propre est une mesure du pourcentage de variance expliqué par ce vecteur propre. Un point important et qui reste une voie de recherche est le choix du paramètre g . En effet, celui-ci détermine la dimensionnalité intrinsèque de l'espace des visages.

Pour la problématique de reconstruction des images de visages, la dimension g de l'espace propre F est estimée pour minimiser l'erreur de reconstruction moyenne ε , calculée comme suit :

$$\varepsilon = \frac{1}{N} \sum_{l=1}^N \left\| A_l - \sum_{i=1}^g (\Phi_i \Phi_i^T A_l) \right\|^2 \quad (5.2)$$

La dimension de l'espace propre est estimée généralement à l'aide du rapport signal / bruit. On définit ce rapport signal / bruit en octave (égal à 3db) par :

$$SNR = \log_2 \left(\frac{\frac{1}{N} \sum_{l=1}^N \|A_l\|^2}{\varepsilon} \right) \quad (5.3)$$

Ce rapport définit une mesure de reconstruction des images ; on estime qu'à partir de 7 octaves la reconstruction de l'image est correcte.

Pour notre problématique de reconnaissance, la détermination de la valeur optimale de g peut être basée sur l'étude du spectre des valeurs propres λ_i (voir figure 5.5). Un algorithme naturel pour déterminer g est de chercher la valeur charnière à partir de laquelle les valeurs propres normalisées tendent vers zéro.

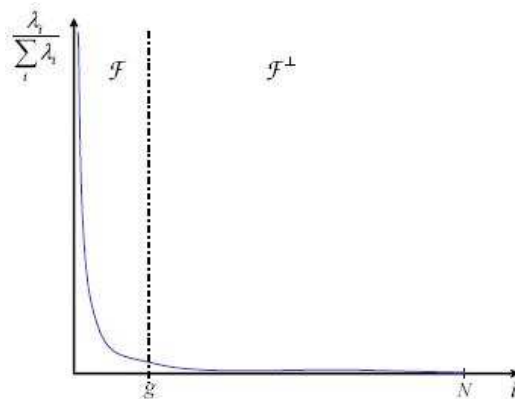


Figure 5.5 : Allure typique des valeurs propres rangées par ordre décroissant.

Dans [TP91], g est défini de manière heuristique à partir de l'étude des valeurs propres. 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. Kirby et Sirovitch [KS90] ont introduit un premier critère de sélection qui est devenu classique par la suite qui est l'énergie ou l'inertie de la dimension [Kir00]. L'inertie de dimension des j premiers vecteurs propres est :

$$\text{inertie} = \frac{\sum_{i=1}^j \lambda_i}{\sum_{i=1}^n \lambda_i} \quad (5.4)$$

où λ_j est la valeur propre associée au $j^{\text{ème}}$ vecteur propre, appelé *Eigenfaces*.

Swets et Weng [SW96] préconisent l'utilisation d'une inertie = 95%, et Kirby [Kir00] utilise un taux de 90%. Kirby [Kir00] a introduit un critère d'étirement, défini comme le ratio $s_i = \lambda_i / \lambda_1$ entre la valeur propre de Φ_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 ($\tau = 1\%$). Ces deux méthodes de sélection offrent à peu près les mêmes performances en terme de taux de reconnaissance.

5.4.1.2 Les *Fisherfaces*

La méthode des *Fisherfaces*, introduite en section 3.5.1.2, est basée sur la résolution de l'Analyse Discriminante Linéaire (*LDA*) selon l'algorithme suivant : la première étape consiste à effectuer une *PCA* dans l'espace original des visages (en retenant M axes principaux), la seconde à projeter la matrice de covariance intra-classe S_w et la matrice de covariance interclasse S_b dans le sous espace principal ainsi défini pour obtenir les matrices de covariances propres S'_w et S'_b , et enfin la troisième étape consiste à construire une *PCA* généralisée sur les matrices de covariances propres S'_w et S'_b . On en déduit la matrice de projection Φ de la *LDA* contenant les g facteurs discriminants.

La technique des *Fisherfaces* ainsi obtenue est dépendante de deux paramètres M et g . Belhumeur et al. [BHK97] proposent de choisir la valeur maximale de M telle que S_w soit de rang plein et la valeur de g suivantes :

$$M = N - k \quad (5.5)$$

$$g = k - 1. \quad (5.6)$$

N et k sont respectivement le nombre d'image et le nombre d'individus de la base d'apprentissage. En appliquant cette technique suivant le protocole d'expérimentation de Lausanne (Expert 1), cela revient à construire une *PCA* sur la base d'images d'apprentissage avec $M = 400$ axes principaux pour y projeter les matrices S_w et S_b . Puis à construire une autre *PCA* généralisée sur les espaces propres de S'_w et S'_b avec $g = 199$ axes principaux.

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 de l'inertie de dimension (voir la section précédente) avec un seuil de 95%. Le paramètre M est donc choisi de manière à ce que le sous-espace principal explique 95% de la variance des vecteurs images, sans tenir compte des besoins de la *LDA* 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 Améliorés de Fisher (MAF). Liu et Wechsler [LW98] ont proposé deux algorithmes dits de Fisher Améliorés :

- La première technique prend en compte les besoins de la *LDA* en aval dans le choix du paramètre M . Leur technique graphique, est basée sur l'étude conjointe du spectre des valeurs propres de la matrice S_T , et du spectre des valeurs propres de la matrice S'_w (un exemple du spectre des valeurs propres est donné en figure 5.5). La matrice de variance totale S_T vérifie la relation suivante : $S_T = S_w + S_b$. On choisit un certain nombre de valeurs candidates M_i pour le paramètre M , dans la région d'inflexion du spectre de S_T . Pour chaque candidat M_i , on trace le spectre des valeurs propres de la matrice correspondante (c'est-à-dire construite dans l'espace de la *PCA* à M_i axes principaux), et on retient le M_i permettant d'obtenir les plus grandes valeurs propres de S'_w . Cette technique de sélection du M optimal est coûteuse et 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'est pas toujours le cas et dépend fortement de la taille et de la variété de la base d'apprentissage.
- Alternativement, Liu et Wechsler proposent d'effectuer la *PCA* préalable avec $M = N - k$, mais par contre de ne retenir que les $M' < N - k$ vecteur 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 du spectre des valeurs propres de S'_w , reste très empirique.

Avec un nombre g suffisant de vecteurs de projection, ces deux algorithmes donnent des taux de reconnaissance comparables sur une sous base de la base FERET. Les résultats obtenus sont meilleurs que ceux de la technique des *Fisherfaces* [SW96]. 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 augmente lorsque g augmente, puis passé g^* les taux de reconnaissance stagnent, voire diminuent. Cependant, ils n'ont pas proposé dans leurs travaux d'heuristiques pour déterminer cette valeur optimale.

5.4.2 Les moments de Zernike

Nous discutons, dans cette section, l'extraction des moments de Zernike comme vecteur caractéristique pour la reconnaissance faciale. Après la normalisation des images faciales, chaque visage est recalculé complètement dans le cercle unité comme le montre la figure 2.3 (b) du deuxième chapitre. En effet, l'approximation qui consiste à calculer partiellement l'image dans le cercle unité réduit considérablement les performances suite notamment à l'exclusion des coins de l'image. Sachant que l'extraction des moments de Zernike ne nécessite pas une phase d'apprentissage, nous avons calculé ces attributs directement sur la base d'enrôlement pour construire les vecteurs caractéristiques des visages de la galerie.

Les moments de Zernike Z_{nm} sont généralement des nombres complexes, comme le montrent les équations 2.1, 2.3 et 2.8 du deuxième chapitre. Nous rappelons que n représente l'ordre de décomposition ($n=0,1,2,\dots,\infty$), et m le nombre de répétitions de la décomposition pour un ordre n donné tel que $|m| \leq n$ et $n - |m|$ soit paire. Cependant, ces nombres complexes sont généralement convertis en plusieurs nombres réels ce qui apporte une certaine robustesse aux variations de luminosités et aux instabilités de normalisation [Ono03]. Les moments de Zernike Z_{nm} peuvent être exprimés comme suit, suite à une conversion non linéaire :

- (1) $\{ \Re(Z_{nm}), \Im(Z_{nm}) \}$: un nombre complexe à une dimension est converti en nombre réel à deux dimensions
- (2) $\|Z_{nm}\|^2$: l'amplitude qui est la valeur absolue du nombre complexe
- (3) $\arg Z_{nm}$: la phase ou l'argument du nombre complexe
- (4) $\{ \|Z_{nm}\|^2, \arg Z_{nm} \}$: l'amplitude et l'argument du nombre complexe

Le tableau 5.4 montre quelques performances d'identification de ces combinaisons de vecteurs caractéristiques que nous avons développés sur la base XM2VTS. Il est clair que le meilleur taux d'identification est réalisé par les moments exprimés sous forme de nombres réels en (1). L'utilisation de l'amplitude (2) ou bien de l'amplitude combinée avec l'argument du nombre complexe (4) comme vecteurs caractéristiques donnent des taux d'identification semblables mais bien inférieur à (1). C'est pourquoi nous exprimons les moments de Zernike dans le reste de notre étude sous la forme (1).

Vecteur caractéristique	(1)	(2)	(3)	(4)
Taille du vecteur	132	66	66	132
Taux d'identification	70.75%	62%	19.25%	62,5%
Intervalles de confiances (\pm)	3,74%	3,99%	3,24%	3,98%

Tableau 5.4 : Taux d'identification pour chaque vecteur caractéristique avec un ordre $n = 10$.

5.4.3 Critères d'extraction des attributs

Le taux d'identification (*Identification Rate - IR*) est la mesure la plus couramment utilisée mais il n'est pas toujours suffisant. En effet, en cas d'erreur, il peut être utile de savoir si le bon choix se trouve parmi les n premières réponses. Nous traçons alors le score cumulé qui représente la probabilité que le bon choix se trouve parmi les n premières réponses. L'allure de la courbe de scores cumulés est d'une importance considérable pour de nombreuses applications qui visent à simplifier l'étape de prise de décision, qui pourrait à la limite être manuelle, sans pour autant chercher à retourner une réponse exacte. Pour ce type de système, nous cherchons à retenir le nombre minimal n de premières réponses, dans lesquelles nous avons la bonne identité avec une probabilité quasi certaine. Pour évaluer ce point de fonctionnement, nous avons relevé le rang à partir duquel nous avons assurément la bonne identité dans l'ensemble de réponses retenu que nous avons baptisé : rang d'identification totale (*Top Matches Rank - TMR*). Ce rang correspond à l'abscisse pour laquelle la courbe de score cumulé atteint les 100% de taux d'identification comme l'illustre la figure 5.6 :

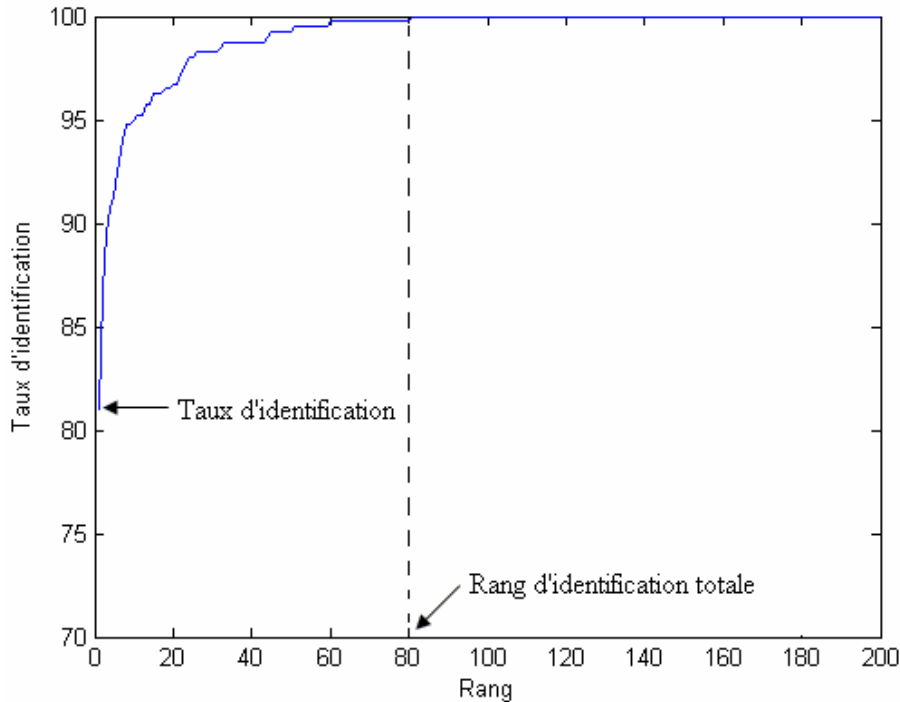


Figure 5.6: illustration du taux d'identification et du rang d'identification totale sur une courbe de scores cumulés.

Le rang d'identification totale (*TMR*) peut donner une indication de la distance maximale que peut avoir une signature requête avec celle lui correspondant dans la galerie. Le choix du vecteur caractéristique qui minimise le rang d'identification totale induit la minimisation de cette distance. Puisque la classification automatique, que nous abordons dans la section suivante, sépare les individus en classes en se basant essentiellement sur la distance entre leurs vecteurs caractéristiques, nous optons principalement pour ce critère (*TMR*) pour choisir les vecteurs caractéristiques qui vont servir de base pour la classification. Nous faisons ainsi la distinction entre les vecteurs caractéristiques extraits pour la classification de la base de données et les vecteurs caractéristiques qui vont servir ensuite pour l'identification. Les vecteurs caractéristiques qui réalisent les meilleurs taux d'identification sont choisis pour accomplir la tâche d'identification. D'autre part, pour la classification et le partitionnement des individus, nous choisissons aussi les vecteurs caractéristiques de taille minimale pour réduire la dimensionnalité du problème de la classification et éviter le fléau de grande dimension (*curse of dimensionality*). Nous pouvons déduire, déjà, un pourcentage de simplification de la base de données (ou galerie) à travers le rapport entre le rang d'identification totale et le nombre d'individus dans la base de données. Ainsi cet indice (*TMR*) est très important pour notre problématique de simplification du processus de l'identification.

5.4.4 Extraction des vecteurs attributs

5.4.4.1 Choix des paramètres *Eigenfaces*

Nous avons évalué les différentes tailles d'espaces de visages pour l'expert 1, relatif au protocole de Lausanne appliqué à la base XM2VTS avec une galerie de 200 personnes. Pour un espace de visage de taille unitaire (un seul axe de projection), l'inertie correspondante à cet axe représente 22% de l'inertie totale de l'espace d'apprentissage. Si nous construisons un espace de visage planaire en retenant les deux premiers axes de projection, leur inertie

cumulée s'élève à 31% de l'inertie totale de l'espace d'apprentissage. Etant donné que la base d'apprentissage est constituée de 600 images de visage, nous avons pu construire des espaces de visages de 600 axes de projections englobant toute l'information d'apprentissage disponible. Nous avons itéré sur la plupart des dimensionnalités et nous avons élaboré le processus d'identification pour chacune. Nous avons ensuite relevé le taux d'identification et le rang d'identification totale pour les différentes tailles possibles des vecteurs caractéristiques. Les courbes des figures 5.7 et 5.8 représentent le taux d'identification et le rang d'identification totale pour les différentes inerties des sous-espaces de visages correspondant à la taille des vecteurs caractéristiques.

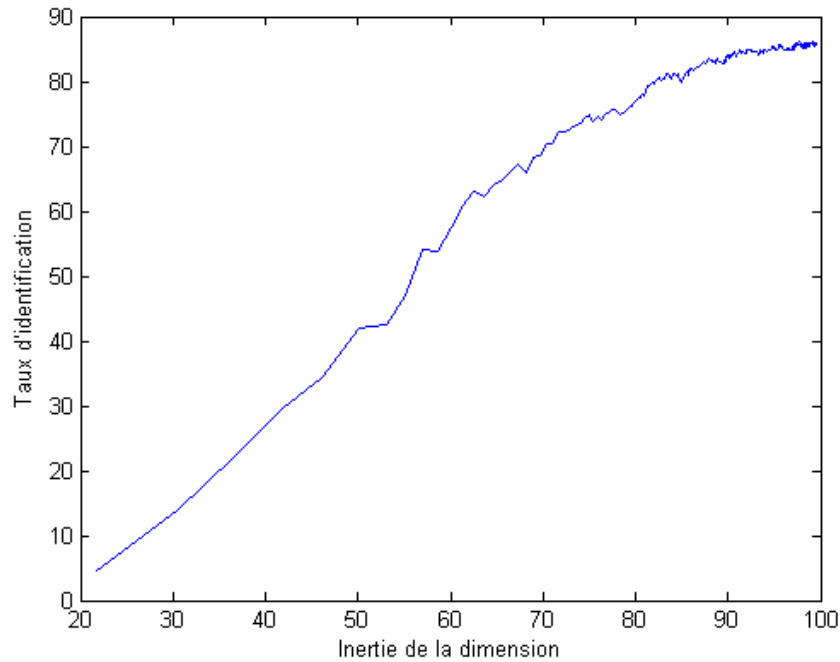


Figure 5.7 : Taux d'identification pour les différentes tailles des vecteurs caractéristiques. Ces différentes tailles correspondent à des inerties allant de 22% à 100% de l'espace de visages.

Nous remarquons d'après la courbe de la figure 5.7 qu'à partir d'une inertie de 90% à 95% du sous espace de visage, nous obtenons un taux d'identification quasi stable de l'ordre de $85\% \pm 3$. Nous validons, ainsi les travaux de Swets et Weng [SW96] et Kirby [Kir00] qui préconisent de retenir 90% et 95% de l'inertie de l'espace total des images. Nous donnons, dans le tableau suivant, à titre indicatif, la dimension correspondante à certaines inerties relevées.

Inertie	50%	60%	70%	75%	80%	85%	90%	95%
dimension	6	11	22	30	44	66	105	190
IR	42%	57,5%	70,5%	74,75%	77,25%	80,5%	84%	85,25%
Int. Conf.(±)	4,05	4,06	3,75	3,57	3,44	3,25	3,01	2,91

Tableau 5.5 : Dimensionnalité de certaines inerties de sous-espaces de visages et les taux d'identification relatifs.

Ainsi, nous retenons ces deux dernières dimensionnalités pour mener les étapes d'identification avec les *Eigenfaces*. D'autre part, nous remarquons, sur la figure 5.8, que le

rang d'identification totale est minimal (autour de 90) pour les inerties des sous espaces allant de 80 à 85% de l'inertie totale de l'espace global des visages. Les vecteurs caractéristiques ainsi extrait sont de tailles allant de 44 à 66. Nous choisissons plutôt le vecteur de taille 44 pour opérer la classification avec les *Eigenfaces* afin de réduire la dimensionnalité de la classification. En retenant par exemple 90 identités sur les 200 de la galerie, nous réalisons un pourcentage de simplification de l'ordre de 55% de la galerie.

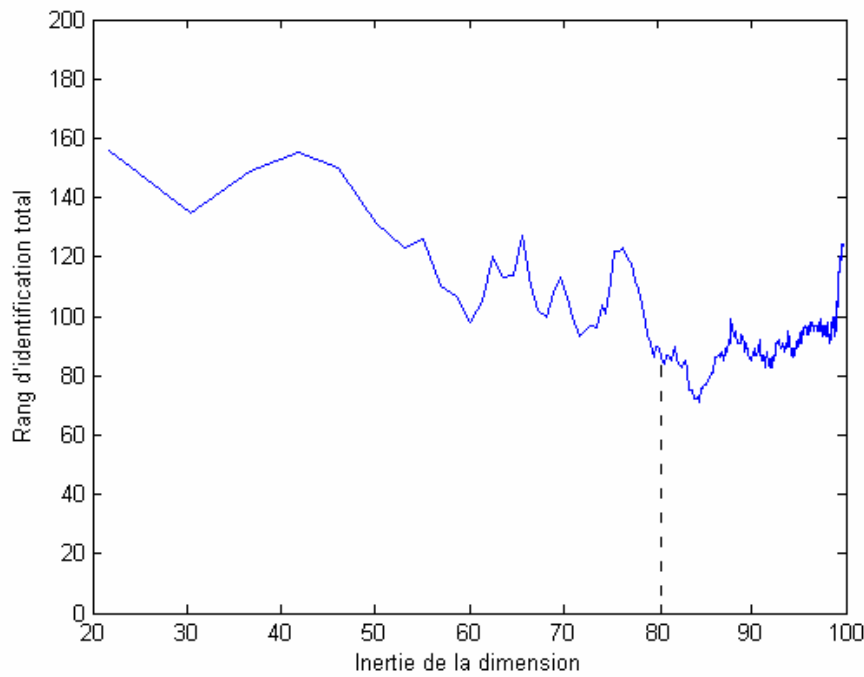


Figure 5.8 : Rang d'identification totale pour les différentes tailles des vecteurs caractéristiques. Ces différentes tailles correspondent à des inerties allant de 22% à 100% de l'espace de visages.

Nous avons étendu ces expérimentations aux experts 2, 3 et 4 définis plus haut avec des bases d'entraînement différentes de celle de l'expert 1. Les résultats obtenus confirment les choix retenus : les inerties de dimension de 80% et de 95% sont optimales pour le choix des vecteurs caractéristiques respectivement pour la classification automatique et pour l'identification.

5.4.4.2 Choix des paramètres *Fisherfaces*

Nous avons, tout d'abord, choisi les paramètres M et g suivant la méthode de Belhumeur et al. [BHK97] qui fixent $M = N - k$, et $g = k - 1$. En appliquant cette méthode suivant le protocole d'expérimentation de Lausanne (Expert 1), cela revient à construire une *PCA* sur la base d'image d'apprentissage avec $M = 600 - 200 = 400$ axes principaux pour y projeter les matrices S_w et S_b , puis à construire une autre *PCA* généralisée sur les matrices de covariances propres S'_w et S'_b avec $g = 199$ axes principaux. Nous avons obtenu les résultats suivants :

M	g	IR	TMR
400	199	88,5% ± 2.62	122 ± 11,34

Tableau 5.6 : résultats d'identification par les *Fisherfaces* avec les paramètres suggérés par Belhumeur et al. [BHK97] sur la base XM2VTS.

Le taux d'identification est nettement meilleur que celui obtenu par les *Eigenfaces*. Par contre, nous obtenons un rang d'identification totale supérieur.

Nous avons étudié, ensuite, les paramètres $M < N - k$ selon la technique de Swets et Weng [SW96]. Le choix du paramètre M étant réalisé à l'aide de l'inertie de dimension. Nous avons essayé différents seuils d'inertie allant de 10% à 90% pour construire le sous espace de visage. Nous avons choisi, par contre, le paramètre g comme étant le minimum de $N - k$ et de M . Le tableau suivant illustre les taux d'identification et le rang d'identification totale pour les valeurs de M correspondantes aux différents seuils d'inertie. Nous remarquons que pour une inertie de 90% de l'espace de visage, M correspond à la valeur de $N - k$ suggéré par Belhumeur et al. [BHK97], avec $N = 600$ et $k = 200$ dans notre expérimentation.

Inertie	90%	85%	80%	75%	70%	65%	60%	55%	50%	40%	35%	30%	25%	20%	15%	10%
M	400	338	284	238	199	165	136	111	90	56	42	31	22	15	9	5
g	199	199	199	199	199	165	136	111	90	56	42	31	22	15	9	5
IR	88,5	88,75	87,25	85,75	88,25	89,25	89	88,75	89	87,5	85,5	82,5	76,5	66,25	55,5	34,5
$Int. Conf.$	3,71	3,67	3,87	4,06	3,74	3,60	3,63	3,67	3,63	3,84	4,09	4,41	4,93	5,50	5,78	5,52
TMR	122	119	155	136	72	71	46	55	52	60	58	49	97	85	114	155
$Int. Conf.$	11,34	11,42	9,71	10,85	11,16	11,13	9,79	10,38	10,20	10,66	10,55	10,01	11,62	11,5	11,51	9,71

Tableau 5.7 : Résultats d'identification *Fisherfaces* pour différents sous-espaces de projection.

Etant donné qu'on cherche à réduire aussi bien le TMR que la dimensionnalité des vecteurs attributs qui vont servir pour la classification, le sous-espace qui renferme 30% de l'inertie totale de la base d'apprentissage de visage optimise l'extraction de ces attributs. Ainsi la signature extraite à partir de cet espace de 31 composantes et conduisant à un TMR de 49 sera préférée à celle de 136 composantes même si elle conduit à un TMR de 49. D'autre part, on remarque qu'à partir de 50% de l'inertie totale des images d'apprentissages les taux d'identification et leurs intervalles de confiances sont quasi identiques. Le sous-espace décrivant 65% de l'inertie totale de la base d'apprentissage conduisant à une signature de 165 composantes semble optimiser ce taux d'identification. Cependant le sous-espace décrivant 50% de l'inertie totale de la base d'apprentissage conduisant à une signature de 90 composantes sera préféré à ce dernier. Nous retenons ainsi les dimensions correspondantes à ces inerties de 30% et 65% pour les vecteurs caractéristiques qui vont servir respectivement à la classification et à l'identification.

5.4.4.3 Choix des paramètres des moments de Zernike

Nous évaluons, dans cette section, la relation entre l'ordre n des moments de Zernike et les indices de performances établis en 5.4.3, à savoir le taux d'identification et le rang d'identification totale. Le tableau 5.8 détaille les moments de Zernike pour les dix premiers ordres. Toutes les combinaisons avec les facteurs de répétitions m positifs ($m \geq 0$) sont illustrées.

Les figures 5.9 et 5.10 montrent la relation entre la taille du vecteur caractéristique qui est relative au nombre des moments de Zernike et leurs performances. Nous avons calculé les

moments jusqu'à l'ordre 40 pour chaque image de visage. Pour un ordre n donné, nous avons calculé la famille des moments d'ordre inférieur ou égal à n avec des facteurs de répétitions positifs pour construire le vecteur caractéristique. En effet, chaque moment ayant un facteur de répétition négatif est redondant avec le moment opposé. Physiquement, ces deux moments correspondent à des parcours opposés de l'image dans le cercle unité.

Ordre	Moments de Zernike
0	$Z_{0,0}$
1	$Z_{1,1}$
2	$Z_{2,0}$ $Z_{2,2}$
3	$Z_{3,1}$ $Z_{3,3}$
4	$Z_{4,0}$ $Z_{4,2}$ $Z_{4,4}$
5	$Z_{5,1}$ $Z_{5,3}$ $Z_{5,5}$
6	$Z_{6,0}$ $Z_{6,2}$ $Z_{6,4}$ $Z_{6,6}$
7	$Z_{7,1}$ $Z_{7,3}$ $Z_{7,5}$ $Z_{7,7}$
8	$Z_{8,0}$ $Z_{8,2}$ $Z_{8,4}$ $Z_{8,6}$ $Z_{8,8}$
9	$Z_{9,1}$ $Z_{9,3}$ $Z_{9,5}$ $Z_{9,7}$ $Z_{9,9}$
10	$Z_{10,0}$ $Z_{10,2}$ $Z_{10,4}$ $Z_{10,6}$ $Z_{10,8}$ $Z_{10,10}$

Tableau 5.8 : Relation entre les moments de Zernike et leurs ordres

La famille des moments d'ordre $\leq n$ contient, en général, $(n + 1) \cdot (n + 2) / 2$ moments. En considérant uniquement les moments de facteurs de répétition positif, le nombre de moments est réduit à :
$$\begin{cases} (n + 2)^2 / 4 & \text{pour } n \text{ pair} \\ (n + 1) \cdot (n + 3) / 4 & \text{pour } n \text{ impair.} \end{cases}$$

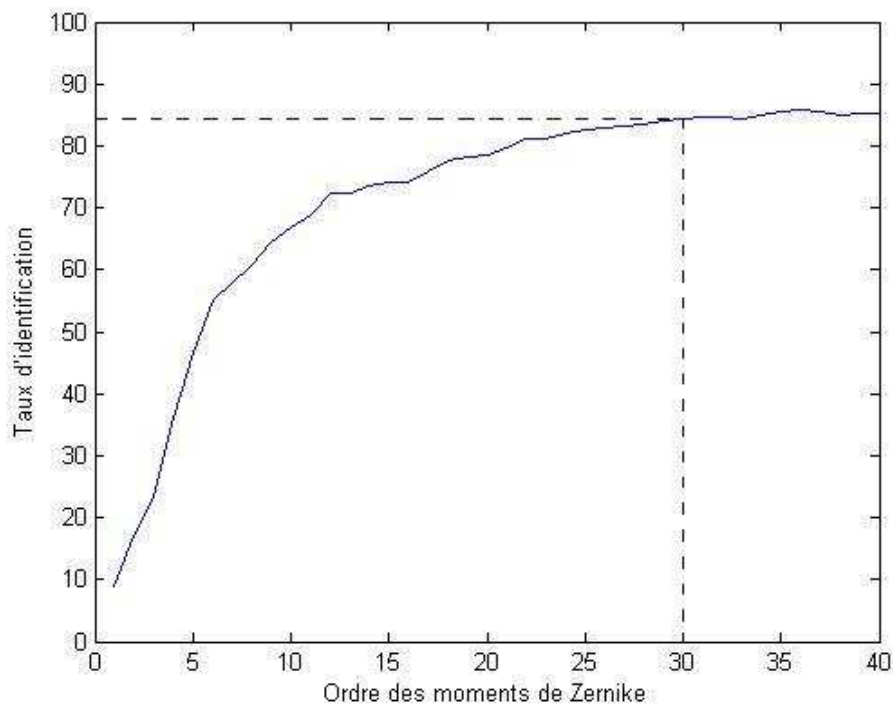


Figure 5.9 : Taux d'identification en fonction de l'ordre des moments de Zernike.

D'autre part, chaque moment est exprimé avec deux composantes : une valeur réelle et une valeur imaginaire. Ainsi le vecteur caractéristique extrait d'une image pour un ordre donné n est de taille $g = \begin{cases} (n + 2)^2 / 2 & \text{pour } n \text{ pair} \\ (n + 1)(n + 3) / 2 & \text{pour } n \text{ impair.} \end{cases}$

Le tableau 5.9 illustre quelques exemples de la taille g des vecteurs caractéristiques pour les ordres n correspondants.

n	1	5	10	15	20	25	30
g	4	24	72	144	242	364	512

Tableau 5.9 : la taille g du vecteur caractéristique correspondant à une famille des moments de Zernike d'ordre n .

Nous remarquons que le taux d'identification se stabilise à partir de la famille de moments d'ordre 30 avec une signature de 512 composantes. Nous obtenons, ainsi un taux d'identification de 84,5% comparable à celui des *Eigenfaces* mais avec un vecteur beaucoup plus grand. D'autre part, en analysant la figure 5.10, le rang d'identification totale atteint ses valeurs minimales à partir de la famille de moments d'ordre 24 ce qui correspond à une signature de 338 composantes. Ce vecteur d'attribut est par conséquent mieux adapté à l'étape de classification qui suivra l'extraction de ces caractéristiques, cependant, leur taille est relativement importante pour l'étape de classification (curse of dimensionality). Un choix plus judicieux de l'ordre des moments de Zernike peut être effectué en évitant une perte conséquente du TMR tout en diminuant au maximum la taille du vecteur attribut. Nous pouvons choisir, par exemple, des signatures aux alentours de 200 composantes avec un ordre n égal à 18.

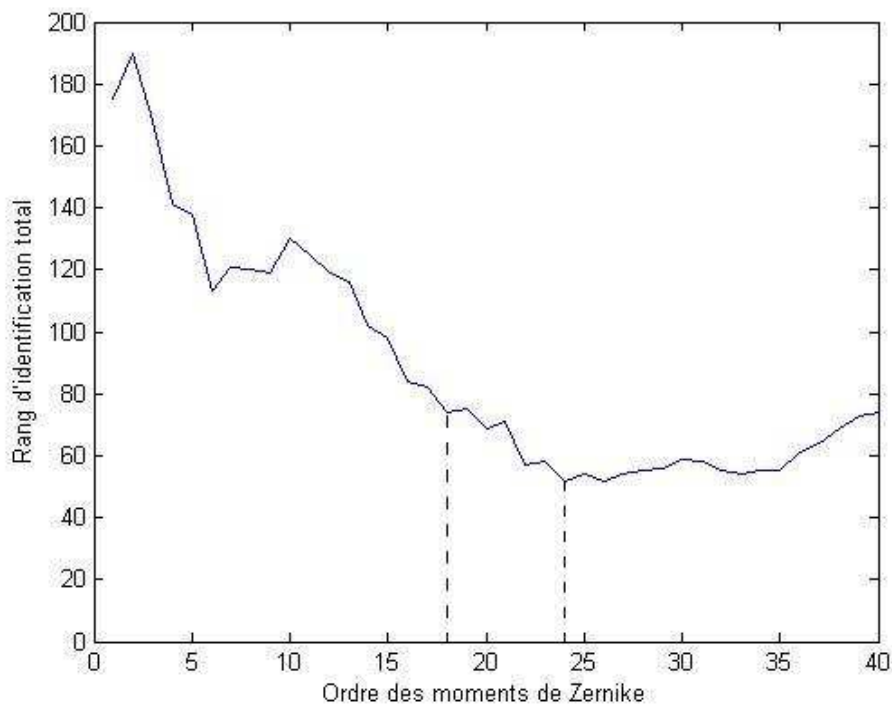


Figure 5.10 : Rang d'identification totale en fonction de l'ordre des moments de Zernike.

Sur la base des résultats de cette expérimentation, il est clair que les moments de Zernike sont moins performants en termes d'identification que les *Eigenfaces* et les *Fisherfaces*. Cependant, pour la tâche de partitionnement de notre base de données, le rang d'identification totale plaide pour un meilleur comportement en termes de classification.

Nous concluons que le choix des attributs et des vecteurs caractéristiques est capital pour optimiser l'opération de la reconnaissance. La précision de la recherche dépend essentiellement de la complétude des données extraites dans le nouvel espace de représentation des visages. Plus les espaces de représentation sont complets, plus les vecteurs caractéristiques décrivent en détail les informations de visages et plus le processus d'identification est performant.

D'autre part, une classe ne décrit pas le détail des caractéristiques de tous les individus qui y appartiennent, mais décrit un mode global pour une moyenne des caractéristiques de ce groupe d'individus. Par conséquent, l'utilisation de vecteurs caractéristiques moins complets (grossiers) que lors de l'identification est optimale quant à la classification. Ainsi, notre critère de rang d'identification totale (*TMR*), nous donne une quantification réaliste de la taille du vecteur caractéristique sur la base duquel la classification serait la plus performante.

5.5 Classification et partitionnement des vecteurs caractéristiques

La morphologie globale des empreintes digitales a été largement étudiée pour partitionner la galerie des identités afin de simplifier la reconnaissance dans cette base de données. Cette opération de simplification de la recherche dans les bases d'empreintes digitales est dite aussi d'indexation puisque les classes sont explicites et connues a priori. Ayant la même perspective de simplifier la reconnaissance, la classification des images de visages suivant le genre ou la catégorie ethnique a été extensivement étudiée. Bien que les classes soient connues à priori, les données extraites des visages sont généralement implicites et aucune forme explicite et mesuré du visage n'a pu être élucidée pour les indexer. Ainsi les attributs extraits des images de visages peuvent être non compatibles avec la distribution des classes apprises et par conséquent non suffisamment discriminantes.

A la différence de ces derniers systèmes qui nécessitent des méthodes de classification supervisé, nous étudions ici la classification automatique pour partitionner les images de visages suivant la similarité de leurs attributs. Cette classification non supervisée a tendance à donner des classes plus compactes et mieux séparées. En effet, la règle générale de la classification non supervisée est de minimiser la variance intra classe des données et de maximiser leur variance inter classes. Nous étudions dans cette section les différentes structures classificatoires pour adresser le problème de partitionnement des bases de données biométriques. Nous étudions ainsi quelques problématiques de la classification automatique afin de développer notre stratégie de partitionnement et de simplification de la recherche dans les bases de données biométriques.

5.5.1 Schéma général de la classification automatique

La classification automatique non supervisée (*clustering*) a pour but de simplifier la représentation des données initiales en réduisant le nuage des points en un ensemble de représentants. La classification automatique se déroule généralement en trois étapes, et à l'aide de quelques paramètres indispensables, à savoir: une mesure de similarité, une structure

de la classification, un algorithme de classification et un critère mesurant le degré d'homogénéité et de séparation des classes.

Pour opérer une classification automatique, nous organisons initialement les données à classifier dans une matrice d'individus \times variables, où les lignes représentent les individus (ou objets) et les colonnes représentent les différentes variables qui les caractérisent. Toute méthode de classification automatique comprend trois étapes [BC07] :

- La première étape du schéma général consiste à transformer le tableau de données initial X qui se présente généralement sous la forme d'une matrice d'individus \times variables en une matrice de distance D ou une matrice de similarité.
- La deuxième étape est la méthode algorithmique de la classification proprement dite. Elle correspond à la transformation de la matrice de distance D en une matrice de distance U représentable par une structure de classification (partition, hiérarchie, arbre, pyramide).
- La dernière étape est la représentation des données, qui infère que chaque structure est liée à une distance appliquée sur la matrice U .

La figure 5.11 illustre les différentes étapes de la classification automatique.

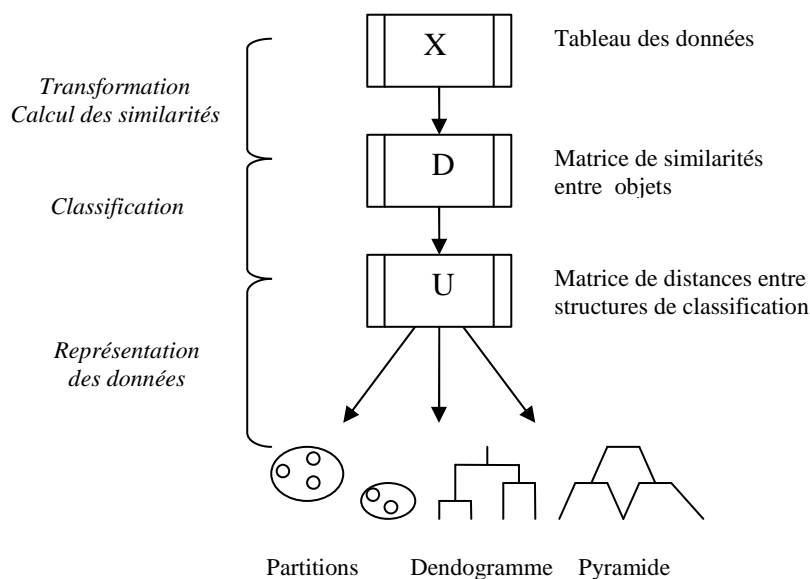


Figure 5.11 : Différentes étapes de la classification automatique, extrait de [BC07].

5.5.2 Structures Classificatoires

Les principales formes que prennent les solutions produites par les méthodes de classification automatique sont une partition ou un recouvrement de l'ensemble de départ, ou encore une collection de partitions ou de recouvrements lorsqu'il s'agit des méthodes hiérarchiques. On parle alors de hiérarchies strictes ou de pyramides. Ces différentes structures sont définies formellement de la manière suivante.

5.5.2.1 Partition

Soit l'ensemble $E = \{e_1, \dots, e_i, \dots, e_N\}$ d'observations, d'individus ou d'objets. Une partition de E est un ensemble $P = \{P_1, \dots, P_i, \dots, P_k\}$ de parties non vides de E d'intersections vides deux à deux et dont la réunion forme E . Ainsi, une partition vérifie les propriétés suivantes :

$$1) \forall l \in \{1, \dots, k\}, P_l \neq \emptyset \quad (5.7)$$

$$2) \forall l, m \in \{1, \dots, k\}, P_l \cap P_m = \emptyset \quad (5.8)$$

$$3) \bigcup_{l=1}^k P_l = E \quad (5.9)$$

Si la propriété 2 n'est pas vérifiée, on parle alors de recouvrement (voir figure 5.12).

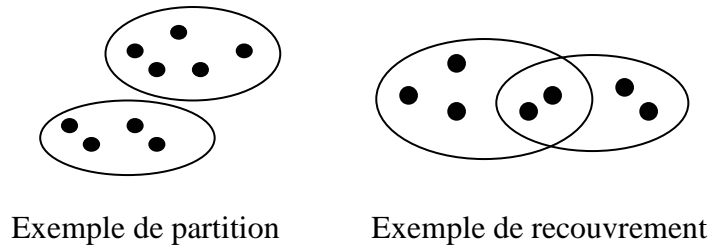


Figure 5.12 : Exemples de structures classificatoires

5.5.2.2 Hiérarchie

Une hiérarchie H est un ensemble de parties de E issues d'une suite de partitions emboîtées. Elle est visualisée par un arbre appelé dendogramme.

Un ensemble H de parties non vides de E est une hiérarchie stricte si :

$$1) E \in H \quad (5.10)$$

$$2) \forall x \in E, \{x\} \in H \quad (5.11)$$

$$3) \forall h, h' \in H \quad h \cap h' \in \{h, h', \emptyset\} \quad (5.12)$$

5.5.2.3 Pyramide

Un ensemble P de parties non vides de E est une pyramide sur E si :

$$1) E \in P \quad (5.13)$$

$$2) \forall x \in E, \{x\} \in P \quad (5.14)$$

$$3) \forall p, p' \in P \quad p \cap p' \neq \emptyset \Rightarrow p \subset p' \quad \text{ou} \quad p' \subset p \quad (5.15)$$

4) Il existe un ordre θ sur E tel que tout élément de P est un intervalle sur E au sens de θ , c'est-à-dire que chaque élément de P possède un minimum et un maximum dans la structure ordonnée (E, θ) .

5.5.3 Mise en œuvre de la classification des visages

La division de la galerie en partitions consiste à regrouper les individus dans des classes homogènes et bien discriminées. Ce regroupement des individus en partitions peut être obtenu avec des structures classificatoires hiérarchiques ou pyramidales. En effet, en employant des structures arborescentes, par exemple, il est possible d'aboutir à un résultat de partition en considérant les feuilles de l'arbre généré (appelé aussi dendogramme) comme des clusters. Cependant, si la visualisation des regroupements sur différents niveaux de l'arborescence générée est souhaitée, il faut faire attention de ne pas tomber dans un résultat d'inclusion. Pour aboutir à un résultat de partitionnement, il faut limiter la troisième propriété de la structure hiérarchique de manière à ce que :

$$\forall h, h' \in H \quad h \cap h' = \emptyset \quad (5.16)$$

Ainsi, nous développons dans le reste du présent chapitre et dans le chapitre suivant les méthodes de clustering débouchant sur des partitions sans recouvrement. Les points de données correspondant aux individus sont assignés aux classes générées C_i de façon à ce que chaque classe contienne au moins un point et que chaque point puisse appartenir à une et une seule classe.

Nous avons mis en œuvre la méthode des k-moyennes, que nous détaillons ci-dessous, pour partitionner la base d'identité suivant un ensemble de variables. Les variables ici correspondent aux différents attributs des signatures ou des vecteurs caractéristiques extraits. Chaque vecteur caractéristique d'un individu est assimilé à un point de données multi dimensionnel occupant le nouvel espace de visage. Ainsi, la taille g des vecteurs caractéristiques extraits constitue la dimensionnalité du problème de classification.

5.5.3.1 Algorithme k-moyennes

L'algorithme de regroupement autour de centres mobiles, généralement imputé à Forgy [For65], est connu en anglais sous le nom de k-means et attribué à MacQueen [Mac67]. Le principe de cet algorithme consiste à construire, à partir d'un ensemble de données, une suite de partitions de manière à minimiser les distances intra-classe ce qui revient à maximiser l'inertie interclasse.

La différence entre les deux approches de MacQueen et Forgy réside dans le calcul des nouveaux centres de gravité (centroïdes) des classes. En effet, l'algorithme de MacQueen recalcule les nouveaux centres dès qu'un objet a été transféré d'une classe à une autre, alors que dans l'algorithme de Forgy, le calcul des centres de gravité est effectué une fois que tous les objets sont réaffectés à la classe la plus proche.

L'algorithme de MacQueen se déroule de la façon suivante :

- Initialisation : Sélectionner les K centroïdes initiaux,
- Affectation : Affecter chaque individu au centroïde le plus proche au sens de la distance euclidienne,
- Représentation : Recalculer les centroïdes,
- Répéter les étapes d'affectation et de représentation jusqu'à la convergence de l'algorithme.

Le principal inconvénient de cet algorithme est la sensibilité aux conditions initiales. En effet, le calcul des centroïdes, après chaque affectation d'un individu, dépend de l'ordre d'affectation des objets. D'autre part, le choix des centroïdes initiaux influence le résultat de la partition finale. Ce choix est souvent effectué d'une manière aléatoire ce qui donne généralement des résultats mitigés [Chr04].

Le choix du nombre de classes K a également une grande influence sur le résultat final. Plusieurs travaux ont été menés pour pallier ces deux inconvénients majeurs de la méthode. L'algorithme X-means [PM00] par exemple, est une extension de k-means permettant de calculer la valeur de K . Dans [BFR98], l'algorithme k-means est amélioré afin de rendre les résultats moins dépendants de l'ordre d'affectation des objets.

5.5.3.2 Initialisation des centroïdes

Afin de partitionner efficacement le nuage de points dans l'espace multi dimensionnel des attributs, nous avons conçu une méthode originale pour choisir les centroïdes initiaux. Etant donné que nous cherchons à construire des partitions qui soient bien séparées, notre premier critère pour initialiser les centroïdes est de maximiser la distance qui les sépare. Nous décrivons ci-dessous les étapes de notre algorithme d'initialisation :

Etape 1. Sélectionner les deux premiers centroïdes ayant la plus grande distance entre eux parmi le nuage de points

Etape 2. Itérer jusqu'à obtenir les K clusters

Etape 2.1. Pour chaque point non sélectionné comme centroïde, calculer la distance qui le sépare de tous les centroïdes déjà sélectionnés.

Etape 2.2. Sauvegarder pour chaque point la distance minimale le séparant du centroïde le plus proche

Etape 2.3. Choisir comme nouveau centroïde parmi tous les points non sélectionnés celui ayant la distance maximale qui a été sauvegardée.

Figure 5.13 : Pseudo code de l'algorithme d'initialisation des centroïdes

Pour mesurer la qualité de la classification automatique, nous avons développé les deux critères usuellement utilisés à savoir la similarité intraclasse qui traduit l'homogénéité ainsi que le degré de séparation interclasses [ZRL96]. Les mesures de l'homogénéité sont la moyenne des rayons W_r et la moyenne des diamètres W_d explicités ci-dessous :

$$W_r = \sum_{l=1}^k \frac{1}{|C_l|} \sum_{x \in C_l} d(x, \mu_l) \quad (5.17)$$

$$W_d = \sum_{l=1}^k \frac{1}{|C_l(C_l - 1)|} \sum_{i=1}^{C_l} \sum_{j=1}^{C_l} d(x_i, x_j) \quad (5.18)$$

Le centroïde d'une classe C_l est donné par : $\mu_l = \frac{1}{|C_l|} \sum_{i=1}^{C_l} x_i$

où $|C_l|$ est le nombre d'individus appartenant à la classe C_l .

Le rayon mesure la distance entre les points de données et le centre de leur classe d'appartenance. Le diamètre est la moyenne de la distance entre tous les points de données d'une classe (voir annexe B). W_r et W_d donnent la moyenne des rayons et des diamètres, respectivement, sur toutes les classes.

D'autre part, le degré de séparation de deux classes est traduit souvent par la distance entre leurs centroïdes. Nous avons calculé cette distance entre les centroïdes de toutes les classes pour en déduire la moyenne de ces distances que nous baptisons B .

$$B = \frac{2}{|k(k-1)|} \sum_{i=2}^k \sum_{j=1}^{i-1} d(\mu_i, \mu_j) \quad (5.19)$$

Notre méthode d'initialisation consiste à rechercher les centroïdes les plus distants possibles. Ceci est réalisé en vue d'arriver à une partition séparant au mieux le nuage de points qui représentent les individus de la galerie. Pour évaluer la qualité de notre initialisation, nous avons comparé son résultat de classification avec une vingtaine de classifications initialisées aléatoirement sur les vecteurs caractéristiques *Eigenfaces*. Le tableau 5.10 explicite les résultats des trois critères d'évaluation présentés ci-dessus. La première colonne comporte les résultats de la partition qui minimise W_r tandis que la deuxième colonne donne les résultats de la partition qui maximise B parmi les vingt partitions obtenues par les initialisations aléatoires. La moyenne et l'écart-type de ces vingt partitions à initialisations aléatoires sont illustrés dans la troisième et la quatrième colonne du tableau.

	20 initialisations aléatoires				Notre initialisation
	Min W_r	Max B	Moyenne	Sigma	
W_r	3,531	3,792	3,840	0,101	3,163
W_d	5,535	6,031	6,084	0,180	5,001
B	8,639	8,639	8,294	0,204	9,548

Tableau 5.10 : Comparaison des critères d'homogénéités et de séparabilité entre les classes calculé par les k-moyennes.

Notre méthode d'initialisation donne une très bonne séparabilité entre les classes générées par la méthode des k-moyennes. En effet, la distance interclasse $B = 9,548$ est supérieure à la distance interclasse maximale de vingt essais avec des initialisations aléatoires des centroïdes. Cette performance est objective vu que l'homogénéité intraclasse mesurée par les deux critères W_r et W_d est meilleure que celle minimisant ces critères parmi les 20 initialisations aléatoires. Les résultats de ces différentes classifications effectuées sur le même nombre de classes $K = 30$, montrent que notre méthode d'initialisation de l'algorithme des K-moyenne implique un bon partitionnement des individus suivant des classes homogènes et très bien séparées. Ces performances ont été également validées pour $K = 50$ et 100 classes.

5.5.4 Résultats de classification

Le taux de classification est la mesure communément utilisée dans les classes pré-établies d'objets ou d'individus. Le taux de classification correspond à la proportion du nombre d'identités correctement trouvées dans la plus proche classe à l'image faciale de l'identité requête par rapport au nombre total des images de test. Nous dénotons par C-Identification ce taux de classification.

5.5.4.1 Partition grossière de la base de recherche

Pour commencer, nous avons partitionné notre galerie en deux puis en trois classes avec la représentation *Eigenfaces*. Pour chaque image de test nous avons calculé les mêmes attributs élaborés pour la partition de l'espace des individus. Chaque image requête de l'ensemble de test se voit affecter la classe ayant le centroïde le plus proche de son vecteur attribut. La probabilité que cette classe d'affectation contienne l'identité recherchée correspond au taux de classification. Le taux de classification (C-Identification) est calculé à partir de l'ensemble des images de test. Le tableau 5.11 présente les résultats suite à une classification en deux et en trois classes de la galerie avec chacune des trois méthodes de représentation développées à savoir les *Eigenfaces*, les *Fisherfaces* et les moments de Zernike.

C-Identification	<i>Eigenfaces</i>	<i>Fisherfaces</i>	Moment de Zernike
K=2	90,75% ± 2.38	91% ± 2.35	90% ± 2.46
K=3	88,50% ± 2.62	89% ± 2.57	88% ± 2.67

Tableau 5.11 : Taux de C-Identification avec un partitionnement en deux et en trois classes.

Les résultats de classification sont similaires pour les différentes méthodes utilisées de représentation de visage. En augmentant le nombre de classes (de 2 à 3), nous remarquons une légère diminution du taux de classification. Cette diminution continue à s'amplifier en augmentant encore le nombre de classes K . Cette diminution est tout à fait prévisible étant donné que la probabilité de trouver la classe recherchée sans connaissance a priori diminue considérablement en augmentant le nombre de classes.

Le taux d'erreur de la recherche *RER* correspond à la probabilité inverse du taux de classification ($RER = 1 - C\text{-Identification}$). Il est évident que la classification en deux partitions minimise ce taux d'erreur. Or, le fait d'obtenir un ordre d'erreur de 10% compromet totalement le processus d'identification qui suivra et diminue considérablement ses performances.

5.5.4.2 Robustesse de la classification

Nous évaluons la robustesse de cette étape de clustering à l'encontre des variabilités intrinsèques qui peuvent affecter les images de visages. Sachant qu'un visage humain peut simuler des milliers d'expressions différentes, la quantification et la mesure des effets de ces expressions sont très difficiles à évaluer dans la reconnaissance biométrique. De plus, il n'existe jusqu'à maintenant aucune base de données de visages simulant différents degrés de sourire ou d'autres expressions faciales. Ces données sont nécessaires pour pouvoir évaluer le degré de robustesse et les limites d'un système de reconnaissance biométrique.

Pour donner une certaine mesure de la limite de notre étape de classification non supervisée vis-à-vis de ces variations dans l'image de visage, nous avons introduit un bruit blanc uniforme. Bien que ce mode de bruitage est loin de modéliser les variations intrinsèques des images de visages, il est facile de le simuler pour parvenir à des changements bien quantifiés dans les images. Ainsi, nous avons simulé plusieurs degrés de bruits blancs sur toutes les images d'enrôlement de l'expert 4' pour constituer une base de généralisation. Ces dernières images de généralisation sont progressivement bruitées comme le montre la figure 5.14.

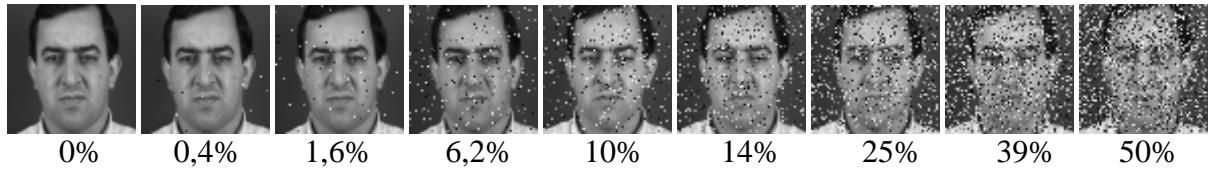


Figure 5.14 : Images de visages bruitées progressivement. Le pourcentage du bruitage figurant est calculé comme étant le nombre des pixels bruités aléatoirement sur le nombre total des pixels de l'image.

La figure 5.15 illustre les résultats d'identification et de classification de ces dernières images de généralisation bruitées progressivement. Il est à remarquer qu'une densité de bruit qui affecte jusqu'à 10% de l'image n'a pas une incidence sur les performances d'identification. Toutefois, on peut aller seulement à 1,6% du degré de bruitage avant de noter la dégradation des résultats de clustering. Au-delà de ce seuil de bruitage, nous enregistrons une baisse significative des taux de classification.

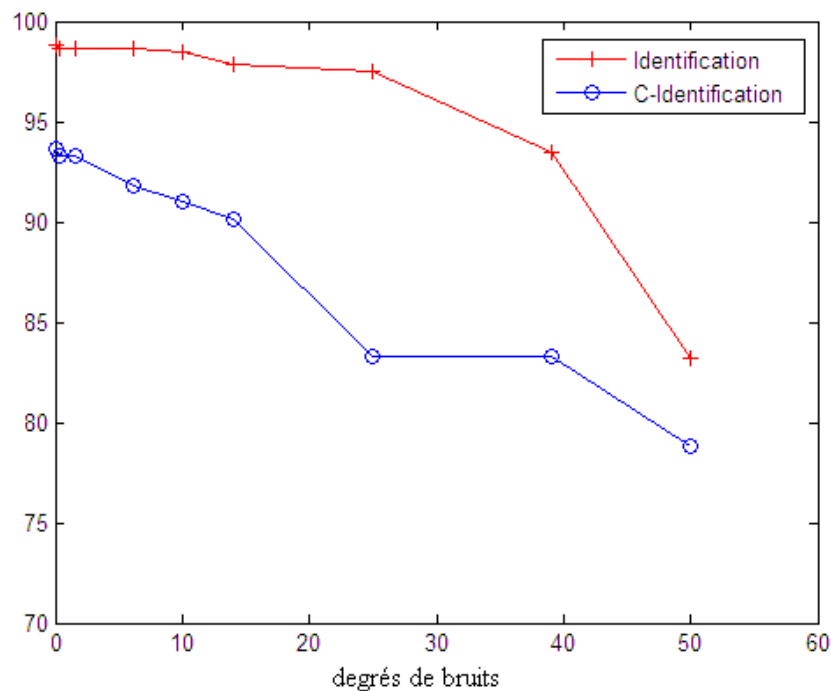


Figure 5.15: Taux d'identification et de classification en fonction du degré de bruitage des images de test.

5.5.4.3 Analyse sémantique des classes

Nous avons conduit une clusterisation en 10 classes des attributs *Eigenfaces* correspondant aux identités de la galerie. Nous visualisons dans la figure 5.16 les quatre premières classes établies, les six autres classes sont données dans l'annexe C.

Nous remarquons que les individus appartenant à chaque classe partagent des traits communs et une apparence similaire. Ces quatre premières classes illustrent bien l'homogénéité et la cohérence entre les individus groupés. La majorité des individus de la deuxième classe partagent des formes du visage, des yeux et une chevelure très similaires. La troisième classe,

d'une cardinalité de 3 individus seulement, contient exclusivement des personnes visiblement d'origine indienne bien que ces personnes forment une minorité dans la galerie. Ce partitionnement de l'ensemble des individus de la galerie témoigne de la force de cet outil de classification non supervisée. La cohérence sémantique des classes obtenues permet d'illustrer l'intérêt de notre approche de classification non supervisée pour partitionner la galerie. Nous illustrons, à droite de chaque classe de visage, les différentes similarités relevées entre ses individus.



Classe n°1

- Couleur de cheveux
- Coiffure
- Genre : féminin



Classe n°2

- Forme de visage
- Couleur de cheveux
- Coiffure
- Genre : essentiellement féminin



Classe n°3

- Origine ethnique
- Moustaches
- Genre : masculin



Classe n°4

- Coiffure
- Front dégagé ou cheveux blonds
- Genre : essentiellement masculin

Figure 5.16 : Représentation des quatre premières classes suite à une clusterisation des vecteurs *Eigenfaces*.

Nous remarquons que les lunettes sont réparties entre les différentes classes. Les attributs portant cette information sont considérés comme secondaires dans l'espace de représentation utilisé.

A l'encontre des trois premières classes qui semblent être fortement homogènes (visuellement), les individus de la quatrième classe peuvent a priori encore être divisés dans des classes plus cohérentes. Pour aboutir à une telle classification, le choix du nombre optimal de classes est la question à résoudre. A cet effet, nous proposons dans ce qui suit une stratégie de clustering guidée par les résultats. Selon la nouvelle stratégie proposée, la classification débouche sur une réduction optimale de la taille de la base de données des identités. C'est une telle classification que nous utiliserons pour partitionner la galerie.

5.5.4.4 Réduction de l'espace de recherche

Etant donné une image faciale dont on veut établir l'identité, son vecteur d'attributs de classification est calculé puis comparé au centre de chaque classe (appelé centroïde). En effet, chaque classe est représentée par son centroïde. L'identité recherchée se trouve potentiellement dans la plus proche classe. Cependant, en retenant pour la suite de l'identification les identités qui se trouvent dans cette partition, l'erreur induite, baptisée taux d'erreur de la recherche (*Retrieval Error Rate - RER*), correspond au rejet de l'identité recherchée par cette étape de simplification. Le taux d'erreur de la recherche compromet les performances de l'identification qui suivra cette étape de simplification. C'est pourquoi, nous faisons le choix de minimiser et d'annuler carrément ce taux d'erreur *RER* pour augmenter les chances de réussite à l'étape d'identification finale. Nous explicitons ci-dessous notre méthode de simplification.

Pour prévenir le rejet ou la simplification de l'identité recherchée de la galerie, nous retenons un nombre P de classes les plus proches de l'identité-requêtes. A l'issue de la classification, en retenant un grand nombre P de classes, le taux de simplification du problème de l'identification se voit marginal et les performances peuvent être difficilement améliorées. Par contre, si on retient une ou relativement peu de classes comme sous-ensemble de recherche pour l'identification, la probabilité d'erreur *RER* va réduire les performances de l'identification qui suivra. L'estimation du nombre P de classes à retenir est assez difficile à évaluer et reste empirique dans le cadre de cette thèse. En effet, le paramètre P est relatif au nombre total de classes K , à la complétude et à l'invariance des vecteurs caractéristiques calculés ainsi qu'à la méthode de classification utilisée. Nous déterminons expérimentalement la valeur de P pour éviter le rejet de la classe contenant l'identité recherché [CLA07, CLA08].

Par analogie aux courbes de scores cumulés (*Cumulative Match Score - CMS*) qui sont utilisées pour l'évaluation des performances de l'identification, nous avons tracé les courbes de classes cumulées baptisées (*BCMS - Binning Cumulative Match Score*). En effet, pour évaluer notre approche de classification, nous avons calculé le taux de classification. Mais en cas d'erreur, il est utile de savoir le rang de la classe dans laquelle se trouve l'identité recherchée. Chaque point d'abscisse x de la courbe de classes cumulées (*BCMS*) représente la probabilité que l'identité requête soit parmi les x plus proches classes. La figure 5.17 illustre deux courbes *BCMS* suite à une classification en 10 et en 30 classes. Nous pouvons estimer le paramètre P des plus proches classes d'une image de test par le rang de classification parfaite pour lequel le taux de classification est de 100%.

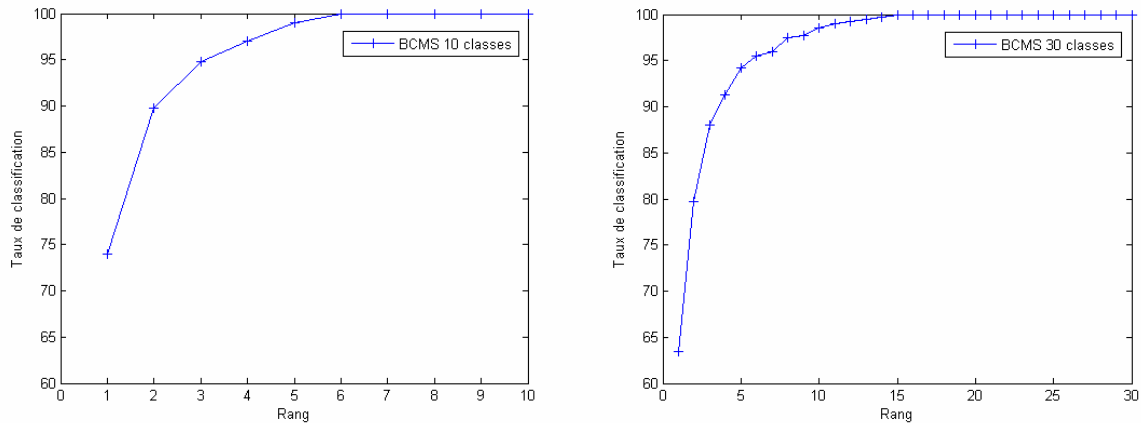


Figure 5.17 : Courbes de classes cumulées (BCMS) avec une classification en 10 et en 30 classes.

Pour une image requête, les plus proches classes forment l'ensemble des identités retenues par notre algorithme pour la suite du processus d'identification. Ainsi, nous assurons un taux d'erreur de la recherche *RER* très petit avoisinant 0%. L'efficacité de cette méthode a été validée en utilisant un autre ensemble de test avec les différentes stratégies d'apprentissage développées au début de ce chapitre.

Le taux de pénétration $R \in [0,1]$ est la fraction de la cardinalité de la partition ainsi retenue par rapport à la cardinalité de la galerie. Le taux de simplification (réduction de l'espace d'identification) du problème d'identification correspondant à la réduction de la taille de la galerie peut être déduit comme étant $1 - R$. Nous remarquons que plus R est petit, plus le taux de réduction de la galerie est grand. En supposant que les classes construites renferment le même nombre d'individus, une valeur approximative du taux de pénétration est donnée comme suit :

$$R \approx \frac{N_B P}{N} \approx \frac{P}{K} \quad (5.20)$$

où N_B est le nombre moyen d'individus par classe, K le nombre de classes et N la taille de la galerie. Cependant, les classes calculées ne contiennent pas toutes le même nombre d'individus. Par exemple, en effectuant une classification sur 10 classes de notre galerie composée de 200 individus, la moyenne de points de données par classe est égale à 20. Mais, en analysant finement ces classes, nous décelons qu'une classe peut comprendre de 3 à 35 individus. Cette inégale répartition est due, sans doute, à l'occupation non homogène des points de données (vecteurs d'attributs) dans l'espace des visages où existent inévitablement des zones denses (surpeuplées) et des zones creuses (éparses). La figure 5.18 trace les taux moyens de pénétration de l'ensemble des images requêtes (400 images de test) en fonction du nombre de classes calculées à partir des vecteurs *Eigenfaces*. Nous traçons aussi les taux minimaux de pénétration obtenus par les images de test réalisant la plus grande réduction de la galerie et les taux maximaux de pénétration obtenus par des images de test réalisant la plus petite réduction de la galerie. Nous remarquons globalement que la simplification est de plus en plus significative lorsque le nombre de classes augmente. La même constatation est faite avec la classification des vecteurs *Fisherfaces* (cf fig. 5.19). Ce comportement est relativement suivi par les moments de Zernike pour lesquels nous obtenons les meilleures réductions avec un nombre allant de 50 à 53 classes.

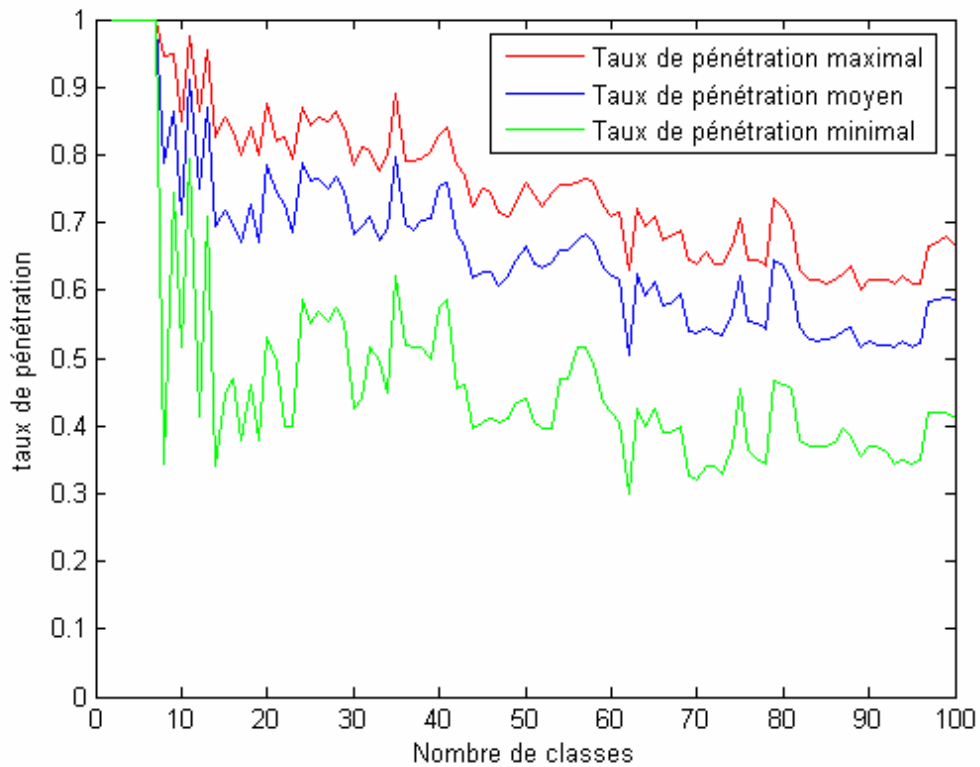


Figure 5.18 : Taux de pénétration maximal, moyen et minimal (parmi 400 images faciales de test) des vecteurs caractéristiques *Eigenfaces* en fonction du nombre de classes

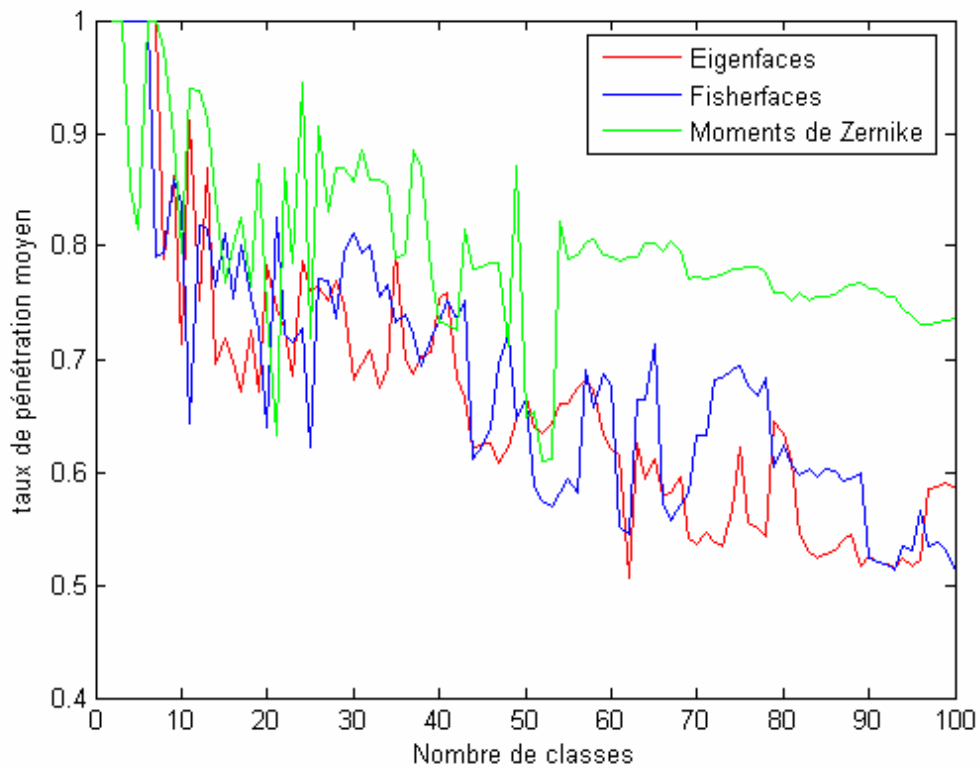


Figure 5.19 : Comparaison des taux de pénétration moyen des *Eigenfaces*, des *Fisherfaces* et des Moments de Zernike.

En augmentant encore le nombre de classes, le comportement du taux de pénétration est confirmé et nous obtenons une plus grande simplification. En abordant la classification avec le nombre maximal de classes, chaque classe est représentée par une seule identité. Le problème de la classification revient à celui de l'identification et le taux de classification correspond à celui de l'identification. Dans ce cas, le nombre P de classes les plus proches de l'image requête et qui contiennent assurément l'identité recherchée correspond au rang d'identification totale TMR comme illustré sur la figure 5.7. Le taux de pénétration est de l'ordre de 30 à 40% et la simplification est assez conséquente.

Cependant cette augmentation du nombre de classes K a un effet pervers sur la complexité temporelle du problème. La classification est réalisée au cours d'une étape hors ligne. Par contre, l'étape d'identification des classes et de réduction de la galerie, qui est en ligne, consiste à comparer le vecteur d'attributs de l'identité requête avec toutes les classes, puis à ordonner par similarité ces classes pour déduire la partition à retenir de la galerie. Cette étape est consommatrice de temps de calcul d'autant plus que le nombre de classes est élevé. Nous traçons ci-dessous la courbe des temps moyens de calcul de l'identification classique avec le temps de notre approche d'identification en fonction du nombre de classes. Le temps d'identification exprimé en secondes correspond au cas de 400 images de test. Nous utilisons ici les vecteurs caractéristiques *Eigenfaces* pour la classification et ceux des *Fisherfaces* pour l'identification. Nous remarquons que plus le nombre de classes augmente, plus le temps de calcul est long et qu'à partir d'un certain nombre de classes, le temps d'identification par notre approche atteint le temps moyen de l'identification classique. Ceci est dû à la complexité croissante de la recherche et de l'ordonnement des classes calculées.

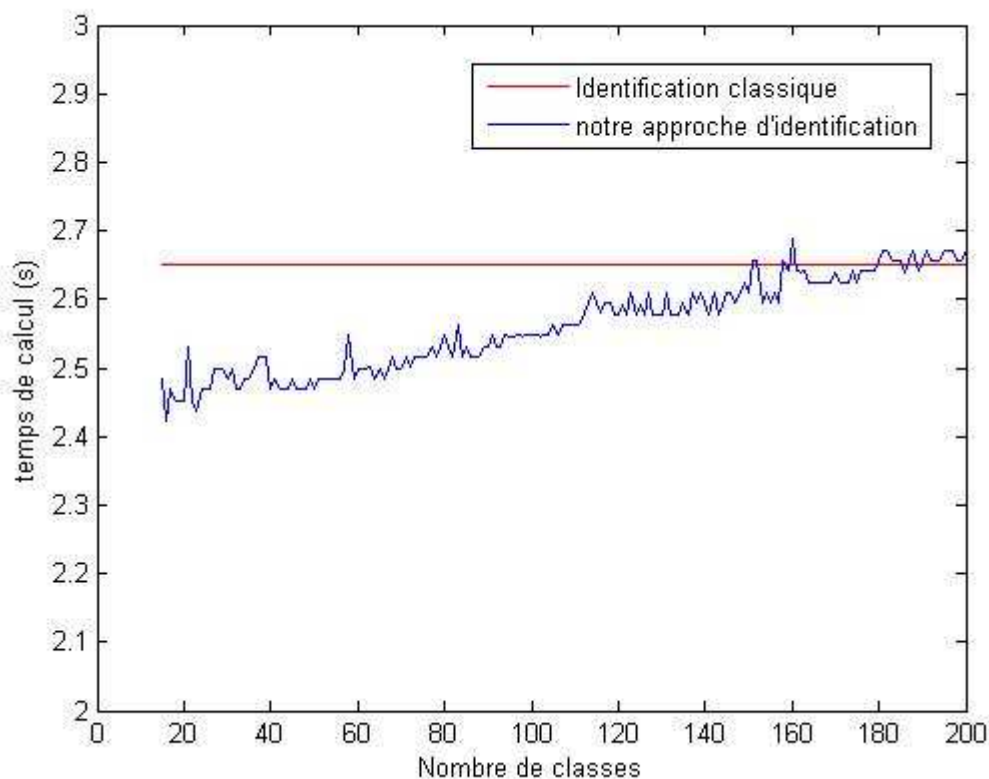


Figure 5.20 : Comparaison du temps de calcul de notre approche d'identification et le temps moyen de l'identification classique en fonction du nombre de classes calculées.

5.6 Combinaison Classification - Identification

Les effets de la réduction de l'espace de recherche sont évidents à partir de cette analyse. Lors d'une identification, les taux de fausse acceptation et de faux rejet intrinsèques aux différentes comparaisons entre l'identité requête et une fraction R tirée au hasard des identités de la galerie sont données par les équations 5.21 et 5.22 :

$$FAR_{R,N} = 1 - (1 - FAR)^{R \cdot N} \approx (R \cdot N) \cdot FAR \quad (5.21)$$

$$FRR_{R,N} = RER + (1 - RER) \cdot FRR \quad (5.22)$$

En opérant un partitionnement et une réduction de la galerie selon notre stratégie, nous arrivons à réduire les fausses acceptations (faux appariements) tout en stabilisant les faux rejets (non appariement). Nous éliminons ainsi toute probabilité d'accumuler les erreurs de la recherche (RER) sur le processus final d'identification. Nous pouvons réécrire le taux de faux rejet FRR comme suit :

$$FRR_{R,N} = FRR \quad (5.23)$$

Le taux de pénétration R de la partition retenue de la galerie assure à la fois la diminution proportionnée de l'erreur de fausse acceptation (appariement) et la conservation du taux de faux rejet (non appariement) en éliminant l'erreur de la recherche RER . Ainsi le taux d'erreur théorique de l'identification sur une telle partition est réduit comme l'illustre l'équation 5.24 par rapport à l'erreur de recherche dans toute la base posé par l'équation 1.9.

$$err \approx 1 - IR \approx R \cdot (N-1) \cdot FAR + FRR \quad (5.24)$$

Il est à remarquer que la mise en œuvre d'une base d'identités de plus en plus grande a pour effet de faire tendre l'erreur induite par notre système d'identification à la proportion R de l'erreur d'un système d'identification classique. Dans ce cas, le taux de faux rejet (FRR) de l'équation 5.24 devient de plus en plus négligeable devant le premier terme.

Supposons que nous ayons un taux d'erreur de 10% dans un système d'identification classique. Si on parvient à réduire de moitié ($R = 50\%$) la base de données grâce à notre stratégie de partitionnement sans rejeter les identités recherchées, la probabilité du taux d'erreur se voit divisée par deux, permettant ainsi d'atteindre 95% comme taux d'identification. Cependant cette amélioration ne peut être obtenue expérimentalement qu'avec l'utilisation de deux méthodes indépendantes l'une pour le partitionnement et l'autre pour l'identification. En effet, plus les vecteurs attributs utilisés pour la classification sont indépendants de ceux utilisés pour l'identification, plus est grande la probabilité d'élaguer et de réduire les identités de la galerie engendrant une fausse identification. De ce fait, l'identification opérée sur le sous-ensemble retenu de la galerie, est plus précise.

5.6.1 Les stratégies d'identification

Nous rappelons, dans le tableau 5.12, pour chaque type de représentation (PCA, LDA et Zernike) la taille des vecteurs attributs retenue pour conduire les étapes de classification et

d'identification. Le nombre optimal de classes K pour effectuer la partition de la galerie avec l'algorithme des K -moyennes est aussi indiqué pour chaque type d'attribut.

	Identification	C-identification	K
<i>Eigenfaces</i>	190	44	70
<i>Fisherfaces</i>	90	31	93
Moments de Zernike	512	200	52

Tableau 5.12 : Tailles des vecteurs attributs et nombre de classes pour chaque type de descripteur.

Neuf stratégies de combinaisons possibles peuvent être exploitées avec ces trois types de descripteurs. Le tableau 5.13 illustre dans ses lignes les différents modes de combinaison que nous avons mis en œuvre pour partitionner puis identifier les identités sur un sous-ensemble de la galerie.

Stratégies	Partitionnement	Identification
1 -	<i>Eigenfaces</i>	<i>Eigenfaces</i>
2 -	<i>Eigenfaces</i>	<i>Fisherfaces</i>
3 -	<i>Eigenfaces</i>	moment de Zernike
4 -	<i>Fisherfaces</i>	<i>Eigenfaces</i>
5 -	<i>Fisherfaces</i>	<i>Fisherfaces</i>
6 -	<i>Fisherfaces</i>	moment de Zernike
7 -	moment de Zernike	<i>Eigenfaces</i>
8 -	moment de Zernike	<i>Fisherfaces</i>
9 -	moment de Zernike	moment de Zernike

Tableau 5.13 : Les différents modes de combinaison des méthodes de représentation des attributs faciaux entre le processus de partitionnement et celui d'identification finale.

Malheureusement, la combinaison des différents descripteurs (*Eigenfaces*, *Fisherfaces* et moments de Zernike) n'améliore pas le taux d'identification obtenu par le schéma classique de l'identification. Mais, cette approche ne dégrade en aucun cas les performances de l'identification classique. Pour examiner le comportement des premières réponses de notre système d'identification, nous avons tracé la courbe de scores cumulés. Les figures 5.21 5.22 et 5.23 illustrent les courbes de scores cumulés des trois premières stratégies du tableau 5.13. Les six autres stratégies fournissent des résultats semblables. Sur chaque figure, nous traçons la courbe de scores cumulés de notre approche d'identification avec des croix et la courbe de l'identification classique avec une ligne continue. Nous avons combiné les différentes méthodes pour la classification et pour l'identification finale. Etant donné que les meilleurs résultats d'identification sont fournis par les *Fisherfaces*, la meilleure stratégie est illustrée par la courbe de la figure 5.22. Cette stratégie est réalisée par la combinaison des attributs *Eigenfaces* pour la classification non supervisée et les attributs *Fisherfaces* pour l'identification.

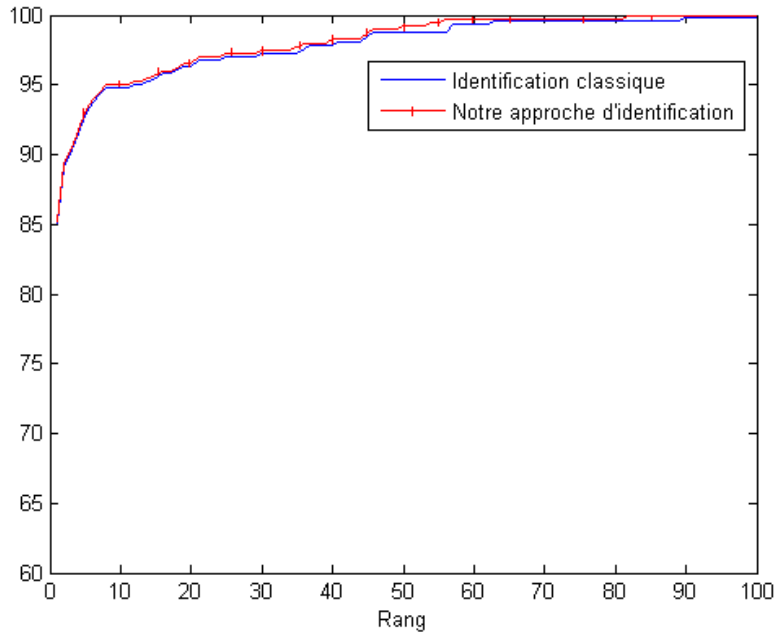


Figure 5.21 : Comparaison des courbes de scores cumulés de notre approche et celle de l'identification classique. La classification est réalisée avec les *Eigenfaces* ($g = 44$) alors que l'identification est réalisée avec les *Eigenfaces* ($g = 190$).

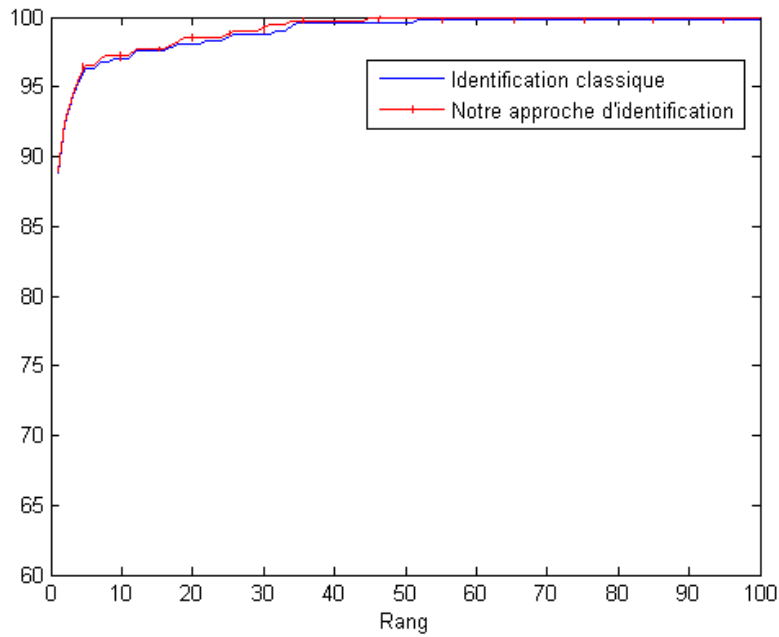


Figure 5.22 : Comparaison des courbes de scores cumulés de notre approche et celle de l'identification classique. La classification est réalisée avec les *Eigenfaces* ($g = 44$) alors que l'identification est réalisée avec les *Fisherfaces* ($g = 90$).

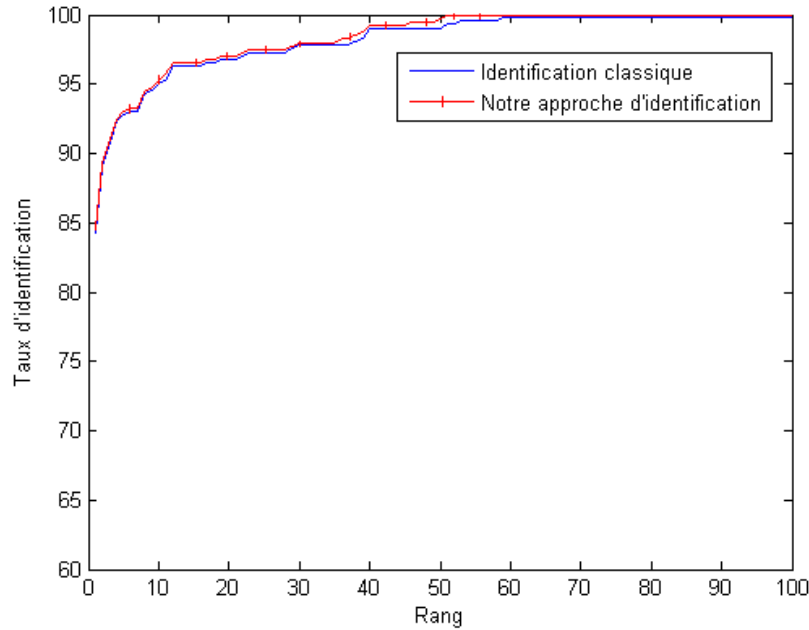


Figure 5.23 : Comparaison des courbes de scores cumulés de notre approche et celle de l'identification classique. La classification est réalisée avec les *Eigenfaces* ($g = 44$) alors que l'identification est réalisée avec les moments de Zernike (ordre = 30).

5.6.2 Coût temporel

Nous comparons, dans ce dernier paragraphe, le temps d'exécution en-ligne de l'identification classique avec celui de notre approche d'identification. Le processus d'identification des partitions (C-Identification) est relativement de faible coût par rapport à l'identification classique. En effet, la C-Identification consiste à calculer seulement la distance des individus aux centres des classes puis à ordonner ces distances en mode croissant, d'autant plus que les vecteurs attributs utilisés pour cette phase sont d'une dimensionnalité réduite par rapport à ceux utilisés pour l'identification (cf. Tableau 5.12). Ainsi nous constatons une baisse notable du temps de calcul pour les stratégies employant les *Eigenfaces* et les *Fisherfaces* pour les phases de classification et d'identification finale des identités. Le tableau 5.14 affiche le temps d'identification des stratégies 1,4 et 7 employant les attributs *Eigenfaces* pour l'identification et les *Eigenfaces*, les *Fisherfaces* et les moments de Zernike respectivement pour la C-Identification. Ces évaluations ont été effectuées sur un PC doté d'un processeur Intel Pentium cadencé à 1,4Ghz et d'une mémoire vive de 256 Mo.

Les stratégies utilisant les moments de Zernike pour effectuer la classification sont handicapées par le temps nécessaire à cette méthode de description pour extraire les attributs de chaque image. Contrairement aux méthodes projectives (*Eigenfaces* et *Fisherfaces*) qui nécessitent une simple projection (produit) de l'image sur l'espace construit, les moments de Zernike extraient les attributs géométriques nécessaires pour la classification à partir de chaque image de test. Ainsi les stratégies employant cette méthode de description que ce soit pour la classification ou pour l'identification finale sont handicapées par un temps de calcul très élevé.

Stratégies	C-Identification	Identification	Temps d'exécution de notre approche	Temps d'exécution de l'approche classique
1	<i>Eigenfaces</i>	<i>Eigenfaces</i>	2,86 s	3,64 s
2	<i>Fisherfaces</i>		1,78 s	
3	<i>Zernike</i>		758 s	

Tableau 5.14 : Comparaison des temps d'exécution en seconde de l'approche classique et de notre approche d'identification pour l'ensemble des 400 images de test.

Cette étude réalisée sur une base de données (XM2VTS) de taille moyenne nous a permis de découvrir le potentiel de notre approche d'identification à la fois en termes de précision et de vitesse d'exécution. L'extension du corpus de travail à l'échelle d'une grande base de données est nécessaire pour évaluer correctement notre approche et étudier l'effet du facteur échelle (scalabilité).

5.7 Conclusion

Nous avons proposé dans ce chapitre une nouvelle approche d'identification biométrique. Nous avons développé l'identification faciale 2D avec les *Eigenfaces*, les *Fisherfaces* et les moments de Zernike. Si notre approche a été ici testée pour l'identification faciale, nous pensons qu'elle reste valable pour toutes les modalités biométriques et pour tous les algorithmes de représentation autres que les trois ici considérés. La phase hors ligne consiste à extraire les différents attributs d'identification et de classification ainsi qu'à partitionner la galerie suivant ces derniers attributs.

Cette dernière phase de classification a pour objectif de réduire l'espace de recherche par élimination des classes les plus dissemblables de l'identité requête. L'identification finale consiste à apparier l'identité-requête avec un sous-ensemble de la galerie résultat de la phase de réduction. Pour éviter le rejet de l'identité recherchée, une stratégie expérimentale est proposée pour bien choisir la partition constituant le sous ensemble retenu de la galerie. Nous avons également étudié plusieurs problématiques relatives à l'apprentissage des données pour l'extraction des signatures, à la taille des données extraites et au nombre de classes calculées.

Nous proposons dans le chapitre suivant d'étendre tout d'abord le corpus d'évaluation à une base de données de plus grande taille obtenue par aggrégation de six bases de visages benchmark. Nous évaluons aussi l'algorithme BIRCH de classification automatique conçu spécialement pour partitionner les bases de données volumineuses. Pour finir, nous testerons quelques méthodes de fusion et de combinaison de classifieurs pour améliorer l'efficacité et la précision de l'identification.

Chapitre VI

Généralisation des données faciales et des méthodes de classification pour notre approche d'identification

6.1 Introduction

Ce chapitre est consacré à la généralisation de notre stratégie en terme de données et en terme de méthodes de classification. Outre les K-moyennes, nous explorons dans ce chapitre la méthode des centres mobiles de Forgy et celle de BIRCH. Ces différentes méthodes de clustering sont appliquées dans le but de partitionner l'espace des visages en des clusters homogènes et bien discriminés. Les trois schémas de description du contenu des images faciales par les attributs *Eigenfaces*, *Fisherfaces* et moments de Zernike sont développés pour aboutir à des signatures variées des identités de la galerie. Après l'analyse de la réduction obtenue par ces méthodes de représentation et de clustering, nous proposons une stratégie de combinaison de classifieurs permettant de valider les données retenues pour l'identification finale.

Ces différentes stratégies d'identification sont évaluées sur un corpus d'images de visages collecté à partir de plusieurs bases de données Benchmark. Ceci est dans le but d'évaluer la portabilité de notre approche sur des données dissemblables ainsi que sa robustesse et son passage à l'échelle (*scalabilité*) face à la taille et à la variabilité de la base de données considérée. Les différentes variabilités d'expression, de poses et de facteur temporel ainsi que l'utilisation d'images de différentes bases acquises par des capteurs différents offrent un cadre d'évaluation d'autant plus réaliste que complexe.

6.2 Approche conceptuelle d'identification biométrique

Notre approche d'identification consiste préalablement à diviser la base de données d'identité en plusieurs ensembles afin de regrouper les identités ayant des caractéristiques similaires et à séparer celles qui ont des caractéristiques dissemblables dans des classes différentes. Nous avons montré, dans la section 5.6, que cette étape de classification permet de rendre plus rapide et plus précise l'étape d'identification finale par l'élagage des partitions les plus dissemblables de l'identité requête. Cependant, cette réduction de l'espace de recherche n'est réellement effective qu'à travers une indépendance accrue entre les différents classifieurs des modules de partitionnement et d'identification finale. En effet, plus le degré de corrélation est élevé entre deux classifieurs différents moins on aura de différences à leurs niveau décisionnel. Par contre, plus l'indépendance des deux classifieurs est assurée plus on aura de chance d'élaguer par le premier classifieur des identités qui engendrent des erreurs d'identification par le deuxième classifieur. Ainsi nous pouvons apporter plus d'efficacité et de précision à l'identification finale des identités.

Dans le but d'assurer une indépendance accrue de différents classifieurs, nous développons, dans ce chapitre, plusieurs configurations de méthodes de représentation et de classification. Tout d'abord, à travers l'ensemble des données collectées, nous avons utilisé plusieurs sous ensembles différents pour l'apprentissage des méthodes de représentation. Ensuite, les trois méthodes de représentation et de description des images faciales, à savoir les *Eigenfaces*, les *Fisherfaces* et les moments de Zernike, sont utilisées pour extraire différents vecteurs d'attributs. Sur la base de ces attributs, les méthodes de clustering à savoir les K-moyennes, les centres mobiles de Forgy et BIRCH sont mises en œuvre pour partitionner l'ensemble des identités de la galerie. Pour finir, différentes mesures de similarité sont considérées au niveau de la classification non supervisée aussi bien qu'au niveau de l'identification finale.

Nous schématisons dans la figure 6.1 l'architecture de notre approche d'identification biométrique ainsi que les différentes méthodes développées. Les méthodes de description des données (*PCA*, *LDA* et Moments de Zernike), de classification non supervisée (K-moyennes, centres mobiles de Forgy et BIRCH) ainsi que différentes mesures de similarité (distance Euclidienne, distance de Manhattan, distance Cosinus) constituent la boîte d'outils que nous avons mis en œuvre durant cette thèse. Nous illustrons, dans un cadre en pointillé, les différentes problématiques relevées pour chaque phase de développement.

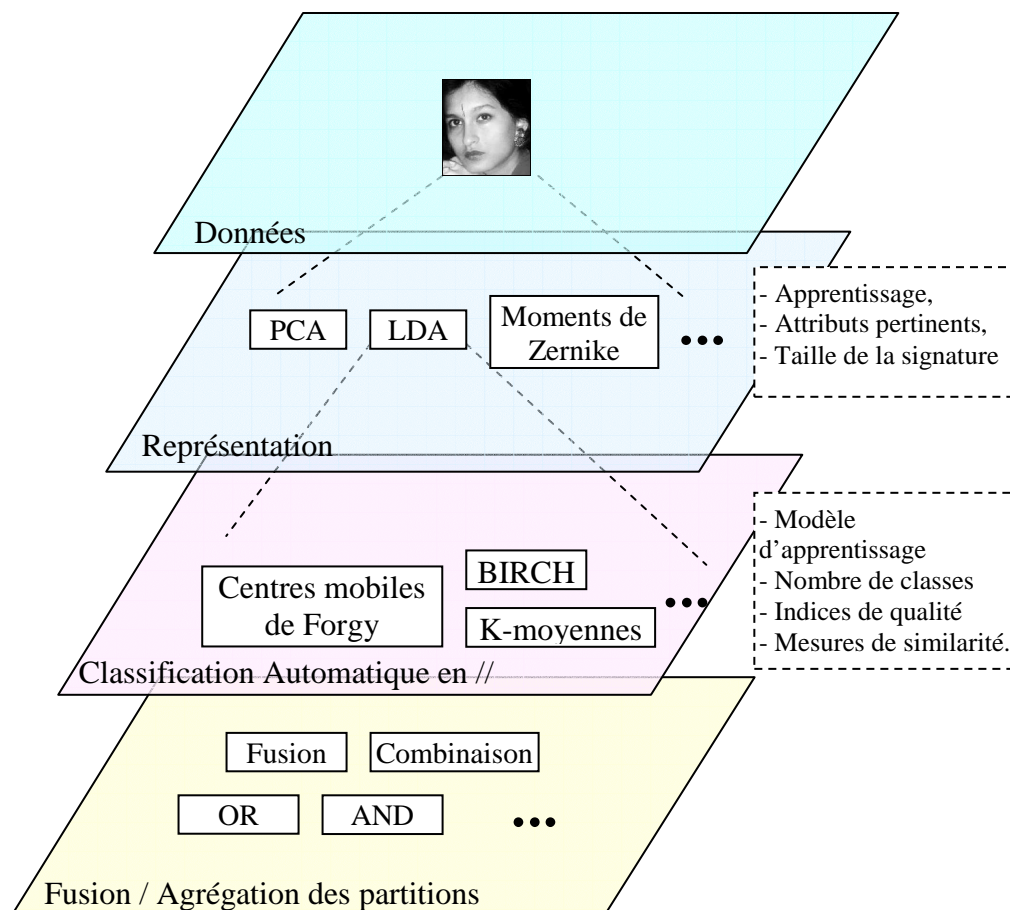


Figure 6.1 : Approche générique d'identification dans les bases de données biométriques

Bien que nous nous soyons limités, dans cette thèse, au développement des techniques ci-dessous, notre approche reste ouverte à toute technique de représentation et méthode de classification différentes. Par exemple, les méthodes bayésiennes des *Eigenfaces* ou la technique *Elastic Bunch Graph Matching (EBGM)* peuvent enrichir la description des visages. Quant aux méthodes de clustering, nous pensons spécialement aux cartes de Kohonen

Nous exploitons, dans cette thèse, uniquement les images faciales 2D pour leur facilité d'acquisition. Cependant, d'autres biométries un peu plus intrusives pouvant être acquises avec une simple collaboration des individus ou extraites sur des scènes de crimes, sont aussi très intéressantes pour les systèmes d'identification biométriques. Citons notamment les empreintes digitales, ainsi que l'iris et les visages 3D pour lesquels les avancées technologiques de capteurs et de techniques d'extraction facilitent continuellement leur acquisition tout en améliorant leur qualité.

Pour "booster" au maximum les performances d'un système d'identification biométrique, une prise en compte du maximum de modalités biométriques pouvant être acquises dans le cadre de l'application est exigée. Chacune des modalités disponibles peut contribuer à la réduction de l'espace de recherche suivant notre stratégie de partitionnement. C'est dans cette perspective que la base de données IV², décrite au chapitre 4, présente un intérêt majeur. Malheureusement, il ne nous a pas été possible de conduire des évaluations dans ce sens. Dans la même perspective, des travaux récents [All09] ont tenté de constituer une base de données chimérique, avec une hypothèse d'indépendance entre les modalités visage, empreinte digitale et forme de la main.

A l'image de l'utilisation de plusieurs modalités biométriques (multi-biométries) pour augmenter les performances de reconnaissance, nous avons utilisé plusieurs autres formes de combinaisons multimodales telles que décrites dans [PJ02]:

- multi-capteurs lorsqu'on associe plusieurs capteurs pour acquérir la même modalité, par exemple des capteurs de différentes résolutions ou de différents type (CCD, CMOS) pour l'acquisition des images de visage.
- multi-instances lorsqu'on associe plusieurs instances de la même biométrie, par exemple l'acquisition de plusieurs images de visage avec des changements de pose, d'expression ou d'éclairage.
- multi-algorithmes lorsque plusieurs algorithmes traitent la même image acquise. Cette multiplicité des méthodes intervient essentiellement dans le module d'extraction des caractéristiques en considérant plusieurs vecteurs attributs et signatures par individus.

Nous nous limitons, pour le développement de ce travail, aux images faciales 2D. Cependant, nous avons pour chaque visage de multiples instances. D'abord nous cherchons pour chaque classe de descripteurs la bonne base d'attributs et la bonne base d'apprentissage (nombre d'instances, choix des instances...). Nous définissons ensuite la base d'enrôlement qui renferme les images des personnes enregistrées. Les signatures de ces images couplées avec une référence unique de chaque identité constituent la base de données biométrique appelée aussi galerie. A partir de cette base d'enrôlement à laquelle on ajoute un bruit blanc, nous pouvons définir une base de généralisation pour tester la robustesse de l'apprentissage (voire section 5.5.4.2).

Nous passons ensuite à la clusterisation. Pour chaque classe de descripteur (*Eigenfaces*, *Fisherfaces* et moments de Zernike), nous choisissons la meilleure base d'attributs et la meilleure méthode de clustering pour partitionner les identités de la galerie.

Nous disposons maintenant des trois partitions de la galerie, les meilleures au vu des trois classes de descripteurs en termes des taux de réduction engendrés. Il s'agit maintenant d'établir l'identification finale des identités sur les clusters retenus par chacun des trois classifieurs. Ainsi, nous explorons un large spectre de combinaison et de possibilités pour optimiser chacune des étapes de notre approche d'identification et pour découvrir les limites de son fonctionnement.

Finalement, il s'agit maintenant d'agréger ces résultats par fusion des classes optimales retenues par chacun des classifieurs. L'identification finale est ainsi effectuée sur les classes et les clusters fusionnés.

6.3 Corpus de données

La réduction de l'espace de recherche pour un système d'identification manipulant une grande base de données d'identités constitue un intérêt majeur pour les applications de surveillance et de sécurité. Comme nous l'avons dit au § 4.5.7, nous avons regroupé une base de travail assez conséquente de 1880 personnes à partir de six bases Benchmark d'images faciales afin d'évaluer les performances (précision, efficacité, robustesse, coût...) de notre approche à l'échelle de telles applications. Le tableau 6.1 rappelle le nombre d'instance (nombre d'images par personne) de chaque base constituant notre corpus de travail.

Base	IV ²	XM2VTS	Color FERET	AR	CVL	ORL
# de personnes	302	295	994	136	113	40
# d'instances	9-14	8	2-22	13-26	3	10

Tableau 6.1 : Nombre de personnes et d'instances constituant chaque sous base de notre corpus de travail.

Nous avons subdivisé l'ensemble des images en trois parties indépendantes : les images d'apprentissage, les images d'enrôlement et les images de test. Par défaut de nombre suffisant d'images par personne dans certaines bases, comme la base CVL qui inclue uniquement 3 images faciales par personne, nous avons développé un ensemble d'enrôlement avec 2 images par personne englobant 3760 images.

Une base de test incluant des images de la majorité des visages de notre base de travail est mise en œuvre avec 3111 images. Ces images de test sont totalement différentes de celles des ensembles d'apprentissage et d'enrôlement.

6.3.1 Choix de l'espace d'apprentissage

Le développement de plusieurs ensembles d'images à partir de la base d'apprentissage nous a permis de construire plusieurs espaces de projection *Eigenfaces* différents. Le nombre et la variabilité des images dans chaque ensemble conditionnent la représentativité de l'espace de projection construit et sa faculté à discriminer les visages. Nous avons fait varier la taille des

ensembles d'apprentissage de 52 à 3671 images. La figure 6.2 représente le taux d'identification en fonction de la taille (en nombre d'images) des ensembles d'apprentissage. L'ensemble d'apprentissage constitué de 1316 images est optimal à notre sens vu qu'il conduit à un taux d'identification de $81,58\% \pm 1,14$. Seul le dernier ensemble d'apprentissage constitué de 3671 images fournit légèrement un meilleur résultat d'identification de $81,68\% \pm 1,14$. La complexité, en termes de temps de calcul, pour construire le nouvel espace *Eigenfaces* requis est de 70 secondes pour le premier ensemble d'apprentissage contre 367 secondes pour le dernier. Vu cette complexité ainsi que l'intervalle de confiance donné, nous avons opté pour le premier ensemble d'apprentissage pour le reste de notre démarche d'identification. Cet ensemble qui renferme deux instances (images de visages) pour chaque individu sélectionné permettra d'entraîner aussi bien l'algorithme *Eigenfaces* et *Fisherfaces* pour la construction de leurs espaces de visages respectif.

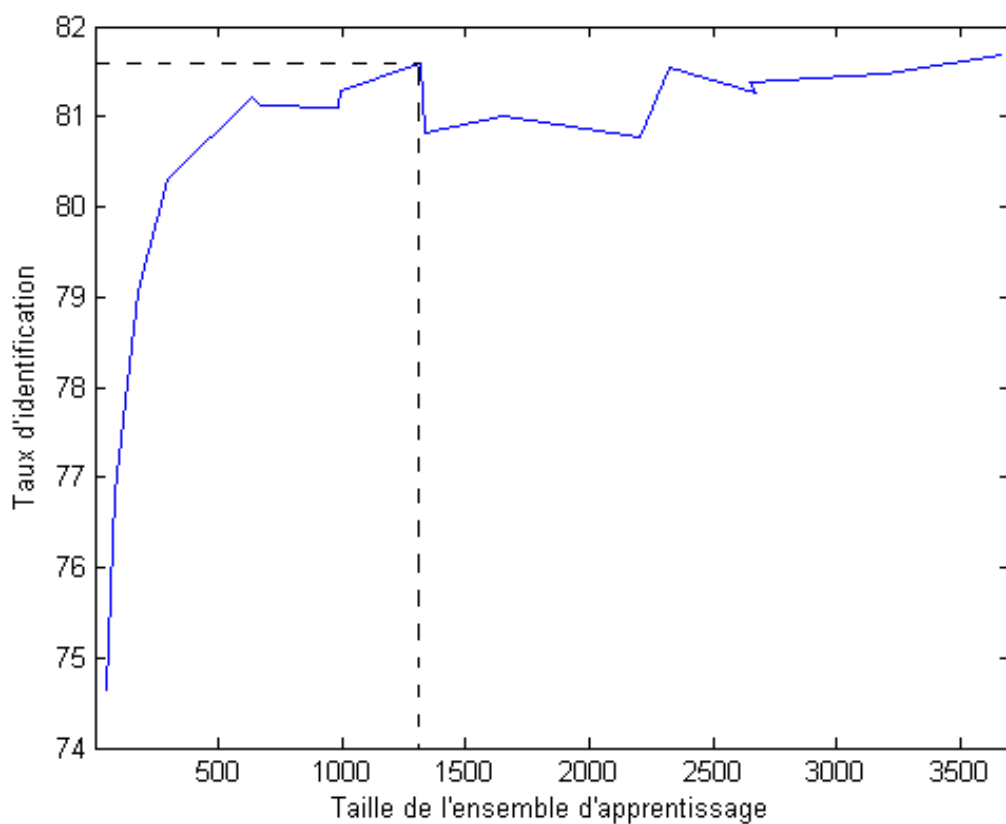


Figure 6.2 : Taux d'identification de la méthode *Eigenfaces* en fonction de la taille de l'ensemble d'apprentissage. Cet ensemble d'apprentissage est requis pour la construction de l'espace de projection pour les méthodes *Eigenfaces* et *Fisherfaces*.

Notons finalement que seuls les ensembles d'apprentissage dépassant les 2300 images renferment quelques instances de la base FERET. Le reste des ensembles d'apprentissage développés renferment uniquement des images des bases de données XM2VTS, IV², ORL, AR et CVL.

6.4 Extraction des vecteurs attributs

Nous présentons, dans cette section, les principales expérimentations de choix des vecteurs attributs à partir de notre grande base de données assemblée. Etant donné la base de visage

constituée, nous avons conduit les mêmes expérimentations que celles de la section 5.4 pour l'extraction des descripteurs *Eigenfaces*, *Fisherfaces* et moments de Zernike. Nous considérons, ci après, les mêmes critères développés dans le chapitre précédent pour mener à bien cette étape d'extraction des caractéristiques, à savoir le taux d'identification (*IR*) et le rang d'identification totale (*TMR*).

6.4.1 Choix des attributs *Eigenfaces*

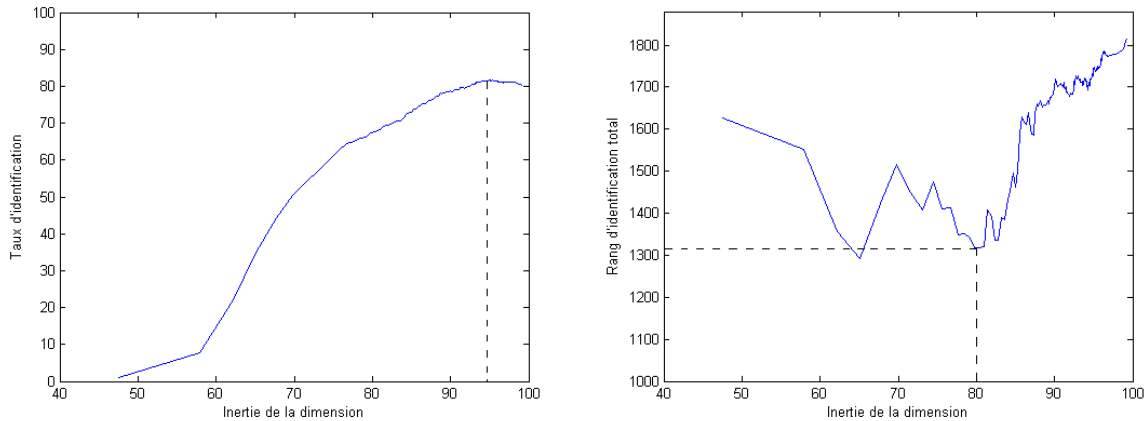


Figure 6.3 : Evolution du taux d'identification (*IR*) et du rang d'identification totale (*TMR*) des *Eigenfaces* en fonction de l'inertie globale du corpus de travail.

D'après les deux courbes de la figure 6.3, le pourcentage d'inertie totale qui maximise le taux d'identification (*IR*) est de 95% alors qu'une inertie autour de 80% de la dimension de l'espace de projection minimise le rang d'identification totale (*TMR*). Ces résultats correspondent exactement à ceux obtenus sur la base XM2VTS, dans la section 5.4. Nous donnons, dans le tableau 6.2, les dimensions des espaces de visages construits relativement à des inerties allant de 60% à 95% de l'espace de visage total.

Inertie	60%	70%	75%	80%	85%	90%	95%
dimension	3	7	10	16	27	54	131

Tableau 6.2 : dimensionnalité de quelques sous espaces de visage relativement à des inerties allant de 60% à 95% de l'espace de visage total.

Pour extraire nos vecteurs attributs *Eigenfaces*, nous retenons les deux espaces de projection renfermant 80% et 95% de l'inertie totale. La dimensionnalité respective de ces espaces notée g est de 16 et 131 axes. Chaque visage Γ_i de la base d'enrôlement dont on soustrait le visage moyen ψ calculé à partir de la base d'apprentissage peut être exprimé comme une combinaison linéaire des axes (appelés aussi vecteurs propres ou *Eigenfaces*) du nouvel espace calculé (équation 6.1).

$$\Gamma_i - \psi = \sum_{i=1}^g w_i \phi_i \quad (6.1)$$

Ainsi chaque image Γ_i de la base d'enrôlement est représenté par le vecteur de poids $\Omega_i = [w_1, w_2, \dots, w_g]$. Le paramètre g est fixé à 16 pour extraire les vecteurs attributs *Eigenfaces* qui vont servir pour la classification non supervisée. Pour mener à bien l'identification, un vecteur attribut plus grand comprenant 131 composantes est extrait des images requêtes.

6.4.2 Choix des attributs *Fisherfaces*

Pour l'algorithme *Fisherfaces*, nous avons étudié plusieurs seuils d'inertie allant de 5 à 85% pour fixer le paramètre M et construire le sous espace de visage selon la technique de Swets et Weng [SW96]. Nous avons ajusté, ensuite, le paramètre g à la valeur de M . Cette valeur de g reste inférieure à la valeur de $k - 1$ suggérée par Belhumeur et al. [BHK97]. En effet, cette valeur n'est atteinte qu'à partir d'une valeur de M correspondant à 90% de l'inertie des composantes principales. Notre base d'apprentissage est composée de $k = 658$ individus avec seulement 2 images par personne, ce qui fait un ensemble de taille $N = 1316$ images.

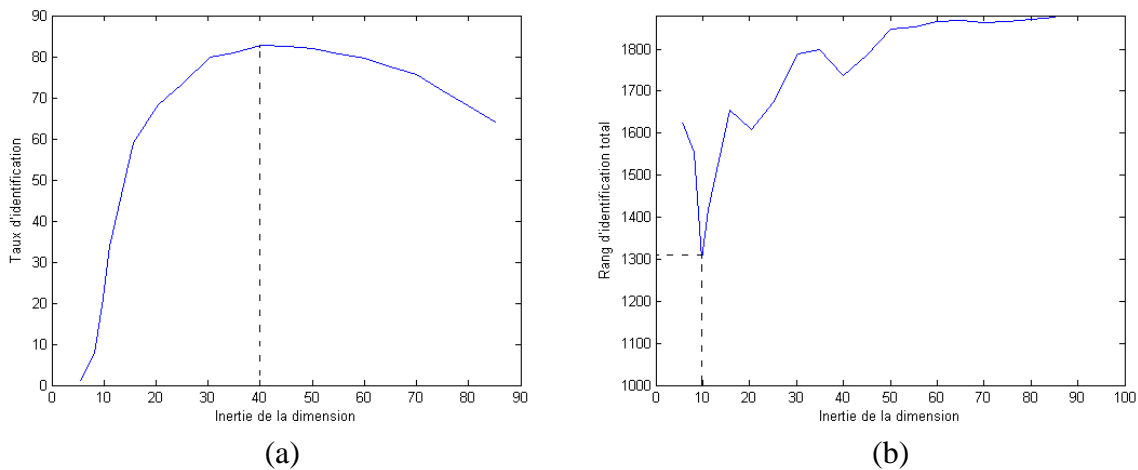


Figure 6.4 : Evolution du taux d'identification (IR) et du rang d'identification totale (TMR) des *Fisherfaces* en fonction de l'inertie de la dimension sur notre corpus de travail.

D'après la courbe de la figure 6.4(a), le pourcentage de l'inertie qui maximise le taux d'identification (*IR*) correspond à 40% de l'espace de visages construit. Ceci correspond à un espace de projection de 67 axes principaux et à un vecteur d'attributs *Fisherfaces* de la même dimension. D'autre part, d'après la courbe de la figure 6.4(b), le rang d'identification totale (*TMR*) est minimisé avec une inertie de 10% de l'espace de visages *Fisherfaces*. Ceci correspond à seulement 4 axes principaux de projection et à un vecteur d'attributs de 4 composantes ($g = 4$). Ce minimum qui semble avoir été obtenu d'une manière accidentelle ne peut être généralisé à d'autres images de test. Cependant en choisissant un vecteur attribut d'une taille voisine de 4, le rang d'identification totale (*TMR*) est bien minimal relativement à l'ensemble de la courbe 6.4b. Ainsi ce *TMR* engendré par le vecteur d'attributs de 4 composantes est loin d'être un minimum local par rapport à ceux des autres vecteurs attributs.

Similairement à la méthode *Eigenfaces*, chaque visage Γ_i de la base d'enrôlement dont on soustrait le visage moyen ψ calculé à partir de la base d'apprentissage peut être exprimé comme une combinaison linéaire des axes principaux, comme le montre l'équation 6.1. Ainsi chaque image Γ_i de la base d'enrôlement est représenté par le vecteur de poids $\Omega_i = [w_1, w_2, \dots, w_g]$. Le paramètre g est fixé à 4 pour extraire les vecteurs attributs *Fisherfaces* qui vont servir pour la classification non supervisée. Un vecteur attribut plus grand de 67 composantes est extrait pour mener à bien l'étape d'identification.

6.4.3 Choix des moments de Zernike

Nous développons ici, la relation entre l'ordre n des moments de Zernike et les indices IR et TMR . Les combinaisons de moments ayant des facteurs de répétition positifs ($m \geq 0$) sont extraites à partir d'une image donnée et forment son vecteur caractéristique. Nous avons évalué les combinaisons de moments jusqu'à l'ordre $n = 20$ pour chaque image de visage.

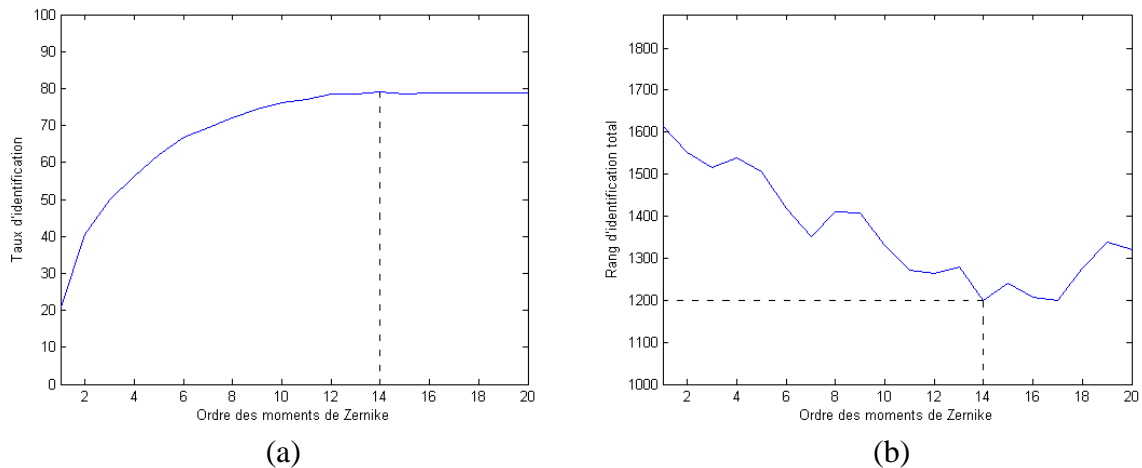


Figure 6.5 : Evolution du taux d'identification (IR) et du rang d'identification totale (TMR) des Moments de Zernike en fonction de l'inertie de la dimension sur notre corpus de travail.

D'après la courbe de la figure 6.5(a), Nous remarquons que le taux d'identification se stabilise à partir de la famille de moments d'ordre 14 impliquant un vecteur d'attributs de 128 composantes. Nous obtenons, ainsi un taux d'identification de l'ordre de 79% comparable à celui des *Eigenfaces* et des *Fisherfaces*. D'autre part, en analysant la figure 6.5(b), le rang d'identification totale atteint ses valeurs minimales à partir de la même famille de moments d'ordre 14. Ainsi, nous pouvons conduire une seule étape d'extraction des moments de Zernike à la fois pour le processus de partitionnement et le processus d'identification.

6.4.4 Synthèse des vecteurs attributs extraits

Nous récapitulons dans ce paragraphe les vecteurs attributs extraits à partir de chaque méthode de représentation des visages. Le tableau 6.3 rappelle la taille g de chaque descripteur compte tenu de son processus d'identification ou de classification.

g	<i>Eigenfaces</i>	<i>Fisherfaces</i>	Zernike
Identification	131	67	128
Clustering	16	4	128

Tableau 6.3 : Taille g de chaque descripteur des trois méthodes de représentation étudiées pour les deux phases d'identification et de clustering.

6.5 Partitionnement de la base de données

Nous présentons dans cette section les algorithmes de clustering choisis pour partitionner et réduire la base enrôlée des images faciales et simplifier par la suite la tâche d'identification.

Les algorithmes de partitionnement, généralement itératifs, permettent de créer, à partir d'un ensemble d'objets, une partition P de K classes C_i avec $i \in [1, K]$. Ces K classes sont telles que :

$$\bullet \quad \forall i \neq j, \quad C_i \cap C_j = \emptyset \quad (6.2)$$

$$\bullet \quad \bigcup_i C_i = P \quad (6.3)$$

Nous pouvons distinguer deux catégories de classifications non supervisées : hiérarchiques et non hiérarchiques.

Dans la **classification hiérarchique (CH)**, les sous-ensembles créés sont emboîtés de manière hiérarchique les uns dans les autres. On distingue la CH *descendante (ou divisive)* qui part de l'ensemble de tous les individus et les fractionne en un certain nombre de sous-ensembles, chaque sous-ensemble étant alors fractionné en un certain nombre de sous-ensembles, et ainsi de suite. Et la CH *ascendante (ou agglomérative)* qui part des individus seuls que l'on regroupe en sous-ensembles, qui sont à leur tour regroupés, et ainsi de suite. Pour déterminer quelles classes on va fusionner, on utilise un critère d'agrégation.

Dans la **classification non hiérarchique**, les individus ne sont pas structurés de manière hiérarchique. Si chaque individu ne fait partie que d'un sous-ensemble, ce sous-ensemble est appelé *partition*.

Méthodologiquement, le partitionnement est relatif à une classification non hiérarchique. Cependant, dans une classification hiérarchique, nous pouvons aboutir à un résultat de partitionnement en visualisant un dendrogramme : arbre hiérarchique indicé. En considérant les feuilles de l'arborescence des données comme des clusters, nous pouvons obtenir un résultat de partitionnement en réorganisant les points de données autour de ces clusters. Ainsi, la méthode *agglomérative* de BIRCH, que nous introduisons ci-dessous, conduit à un résultat de partitionnement.

Nous présentons ci-dessous deux autres méthodes de classification non supervisée qui aboutissent à un résultat de partitionnement : la méthode des centres mobiles de Forgy et la méthode BIRCH. Les résultats de ces deux méthodes ainsi que ceux des K-moyennes sont ensuite détaillés.

6.5.1 Les centres mobiles de Forgy

La technique des centres mobiles de Forgy [For65] est la première variante développée des K-moyennes. Le principe de cet algorithme consiste à construire à partir d'un ensemble de données une suite de partitions de manière à minimiser les distances intra-classe ce qui revient à maximiser l'inertie inter-classe. Comme nous l'avons dit au § 5.5.3, la différence entre les deux approches de MacQueen [Mac67] et Forgy réside dans le calcul des nouveaux centres de gravité (centroïdes) des classes. En effet, l'algorithme de MacQueen recalcule les nouveaux centres dès qu'un objet a été transféré d'une classe à une autre, alors que dans l'algorithme de Forgy, le calcul des centres de gravité est effectué une fois que tous les objets sont réaffectés à la classe la plus proche.

Une des propriétés de cet algorithme est que chaque partition obtenue (i.e. pour chaque itération) est indépendante de l'ordre dans lequel les objets sont affectés. Cette propriété est

liée au fait que les centres des classes sont calculés une fois par itération et que, pour chaque itération, tous les objets sont affectés. Une autre propriété de cette méthode est sa faible sensibilité au choix de la distance utilisée [Sch97] et son adaptabilité au traitement de grandes populations vu qu'on ne calcule pas à chaque étape toutes les distances entre toutes les paires d'individus, mais seulement les distances des individus aux centres des classes.

Cependant, cette méthode présente un inconvénient. La partition obtenue dépend du choix initial des noyaux. Pour pallier ce problème, nous avons repris notre méthode d'initialisation développée dans le chapitre précédent pour séparer au maximum les centres initiaux. Ceci est réalisé dans le but de construire une partition qui sépare le mieux possible le nuage de points correspondant aux individus de la galerie.

6.5.2 La méthode BIRCH

BIRCH (*Balanced Iterative Reducing and Clustering using Hierarchies*) est une méthode de classification conçue par Zhang et al. en 1996 [ZRL96]. Elle permet de réduire l'ensemble initial de données en un ensemble de sous-clusters (sous-classes), présenté sous une forme arborescente, afin de simplifier le problème de classification. Nous nous sommes intéressés à l'algorithme BIRCH conçu initialement pour les bases de données volumineuses et qui a montré de meilleures performances de classification que les méthodes CLARANS et K-moyennes [ZRL97]. Nous présentons les deux notions principales pour cet algorithme : *Clustering Feature* et *CF-tree*. L'approche BIRCH est ensuite détaillée avant de présenter l'algorithme développé pour le partitionnement de la base d'identité.

6.5.2.1 Le *Clustering Feature* : *CF*

Un *Clustering Feature* représente le résumé des informations permettant de définir un cluster donné. Soit une classe C constituée de N points de données d -dimensionnels \mathbf{X}_i où $i = 1, 2, \dots, N$. Le *Clustering Feature* de la classe C est défini par le triplet (N, \mathbf{LS}, SS) , où N est le nombre de points de données dans la classe, \mathbf{LS} est la somme linéaire des N points de données, et SS est la somme au carré des N points de données :

$$\overrightarrow{LS} = \sum_{i=1}^N \vec{X}_i \quad (6.4)$$

$$SS = \sum_{i=1}^N \vec{X}_i^2 \quad (6.5)$$

Théorème d'additivité : On suppose que $CF_1 = (N_1, \mathbf{LS}_1, SS_1)$ et $CF_2 = (N_2, \mathbf{LS}_2, SS_2)$ sont les *Clustering Features* de deux sous-clusters disjoints C_1 et C_2 . Le *Clustering Feature* du sous-cluster formé par la fusion de C_1 et C_2 s'écrit :

$$CF_1 + CF_2 = (N_1 + N_2, \mathbf{LS}_1 + \mathbf{LS}_2, SS_1 + SS_2) \quad (6.6)$$

6.5.2.2 L'arbre CF (*CF-Tree*)

Un arbre *CF* ou *CF-tree* est un arbre à deux paramètres: facteur bifurquant (B pour les nœuds non-feuille « nonleaf-node » et L pour les nœuds feuilles « leaf-node ») et un seuil T . Un nœud non-feuille contient au maximum B entrées de la forme $[CF_i, fils_i]$, où $i = 1, 2, \dots, B$, $fils_i$ son $i^{\text{ème}}$ nœud-fils, et CF_i l'entrée *CF* du sous-cluster représenté par ce fils. Ainsi un nœud

représente un cluster formé par tous les sous-clusters représentés par ses entrées. Un nœud feuille représente un sous-cluster (micro-classe) et contient au maximum L entrées. En plus, chaque nœud feuille possède deux pointeurs : *prev* et *next*, utilisés pour lier tous les nœuds-feuilles ensemble pour un parcours efficace de l'arbre. Mais toutes les entrées d'un nœud-feuille doivent avoir un diamètre (ou bien un rayon) inférieur au seuil d'absorption T .

La taille de l'arbre CF est une fonction de T . Plus T est élevé, moins l'arbre est grand et moins est important le nombre de micro-classes générées. Une fois que la dimension g de l'espace de données est fournie, les tailles des entrées des nœuds feuilles et nœuds non feuilles sont connues, et B et L sont déterminés en fonction de la taille de la mémoire disponible [ZRL96].

L'arbre CF est mis à jour au fur et à mesure que de nouveaux points de données sont insérés. Il est utilisé pour guider l'insertion d'une nouvelle donnée dans le sous cluster approprié. Toutefois, l'arbre CF est une représentation très compacte de l'ensemble de données puisque chaque entrée dans un nœud-feuille n'est pas un unique point de données mais un sous cluster qui engendre autant de points de données que la valeur seuil T le permet.

L'algorithme d'insertion est donné en détail en annexe D.

6.5.2.3 L'approche algorithmique BIRCH

BIRCH s'adresse principalement aux bases de données volumineuses en générant d'abord un résumé compact qui garde autant d'informations que possible sur la distribution de données et en classifiant ensuite le résumé généré, au lieu de la base de données originale.

En termes de temps d'exécution, d'usage de la mémoire, de qualité de la classification et de stabilité de l'algorithme, et en le comparant à d'autres algorithmes, les auteurs [ZRL97] ont présenté BIRCH comme le meilleur moyen de classification disponible qui puisse s'adapter aux larges bases de données. En se basant sur ces études, nous avons choisi l'algorithme BIRCH pour classifier les caractéristiques biométriques extraites à partir des images faciales. La figure 6.6 représente une vue d'ensemble de l'algorithme BIRCH.

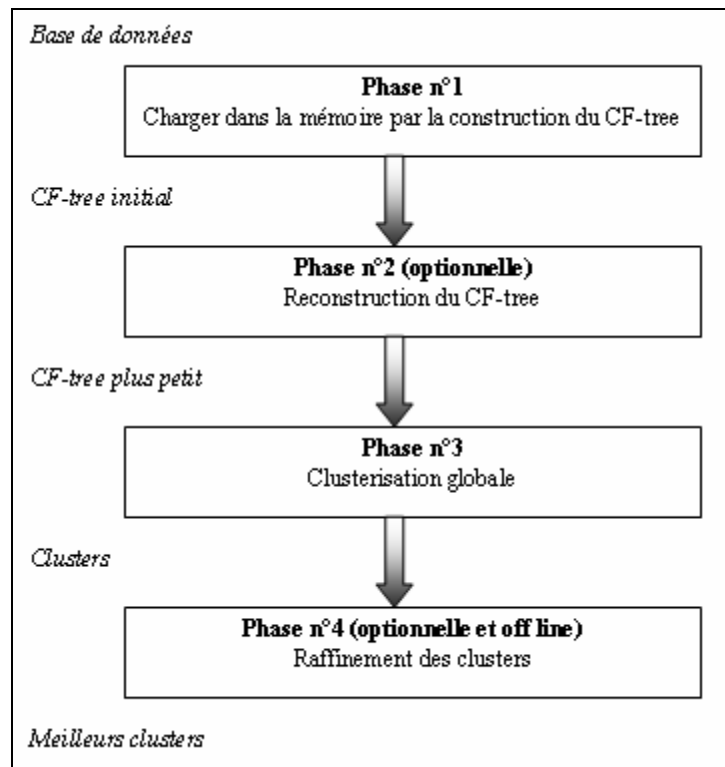


Figure 6.6 : Algorithme BIRCH

Cet algorithme se compose essentiellement de quatre phases :

1. Chargement des données et construction du *CF-Tree* ;
2. Compression des données chargées par reconstruction du *CF-Tree* ;
3. Clusterisation globale des données compressées ;
4. Raffinement des clusters résultants de la troisième phase.

Le rôle de la première phase est la construction d'un *CF-Tree* qui donne le maximum d'informations possible sur les données de départ et que l'espace mémoire disponible permet. Une fois toutes les données allouées dans l'arbre *CF*, la deuxième phase consiste à augmenter la valeur de seuil. Un nouvel arbre *CF* est construit et ce, par réinsertion des entrées de l'ancien arbre dans le nouvel arbre qui s'avère nécessairement plus petit. La troisième phase consiste à examiner les entrées de l'arbre *CF* ainsi construit afin de construire un autre arbre plus petit par élimination des données aberrantes et regroupement des sous clusters les plus empilés dans d'autres de plus grandes tailles. A la suite de cette troisième phase, quelques impuretés peuvent exister à cause d'un mauvais placement des données ou à cause de la troisième phase qui ne s'applique que sur un sommaire grossier de la base initiale. Afin de corriger ces inexacitudes et raffiner davantage la clusterisation, l'algorithme BIRCH intervient par la quatrième phase. Celle-ci utilise les centroïdes des classes résultants de la troisième phase et réorganise les données en associant chaque point de données au centroïde le plus proche.

6.5.2.4 Notre implémentation

Nous avons développé une implémentation de BIRCH en C++. Celle-ci était l'enchaînement du travail de Dorsaf Cherif [Che05] au laboratoire RIADI. Une description détaillée des

différentes classes, des arguments d'entrée et des opérations principales est donnée à l'annexe E. Nous avons développé essentiellement la première phase de construction de l'arbre *CF-Tree* suivant l'algorithme d'insertion décrit dans l'annexe D. Selon les auteurs, les facteurs bifurquant B et L , correspondant à la taille des nœuds et des feuilles respectivement, ont une certaine limite relative à la quantité des données manipulées et à la taille de la mémoire virtuelle utilisée. Notre machine de test, disposant d'une mémoire vive de 1Go et 1,5Go de mémoire virtuelle, permet d'atteindre jusqu'à 300 entrées par nœud pour les vecteurs d'attributs de 128 composantes. Ceci nous a permis de construire des classes contenant jusqu'à 300 individus avec un seuil d'absorption T nul et des classes de plus en plus volumineuses en augmentant le seuil T . Ceci dit, nous avons plutôt besoin de construire des classes de tailles moyennes à petites pour construire des micros classes et développer notre approche de simplification.

Afin de raffiner davantage les partitions obtenues, nous avons adopté, à la suite de cette étape de construction de l'arbre *CF*, la quatrième phase de l'algorithme BIRCH. Celle-ci utilise les centroïdes des classes résultants et réorganise les données en associant chaque point de données au centroïde le plus proche.

Un bon choix de la valeur du seuil d'absorption T peut améliorer considérablement la qualité de la classification générée. Dans [ZRL97], l'estimation de la valeur T a été guidée essentiellement pour optimiser l'utilisation de la mémoire disponible. Les auteurs suggèrent l'utilisation d'une valeur initiale T_0 nulle et l'augmente dynamiquement lors des différentes phases de l'algorithme. D'une façon similaire, nous avons fixé, initialement, le seuil T à une valeur nulle pour la construction de l'arbre *CF*. Dans ce cas, chaque entrée correspond à un seul point de données et une feuille (qui correspond à une micro classe) ne peut renfermer qu'un nombre maximal de L points de données. Nous avons augmenté, ensuite, la valeur du seuil T de façon qu'une entrée puisse absorber d'autres points de données proches de son vecteur *CF* sans ajouter de nouveaux nœuds dans l'arbre généré.

6.5.3 Performances des différentes méthodes de classification

Nous présentons dans cette section les résultats de classification et de réduction de la galerie par les trois méthodes de classification automatique développées ici à savoir : les K-moyennes, les centres mobiles de Forgy et BIRCH. Notre approche de partitionnement a comme premier objectif la réduction de la grande base de données afin de simplifier la complexité de l'identification biométrique. Les résultats sont présentés en termes de taux de pénétration R qui indique en quelque sorte le pourcentage de la partition retenue de la galerie. Bien entendu, on cherche à travers l'étude portant sur le nombre de classes générées par chaque classifieur à optimiser le processus de partitionnement et à minimiser par voie de conséquence le taux R .

6.5.3.1 Réduction de la galerie par l'algorithme des K-moyennes

L'algorithme de classification K-moyennes est initialisé par notre méthode développée dans le chapitre précédent. Au terme de la classification, nous retenons les P plus proches classes parmi les K classes établies. Nous cherchons donc à déterminer expérimentalement la valeur de P pour éviter le rejet de la classe contenant l'identité recherchée. Ceci assure une certaine réduction de la taille de la galerie tout en contrôlant un taux d'erreur de la recherche *NER* égal ou très proche de zéro.

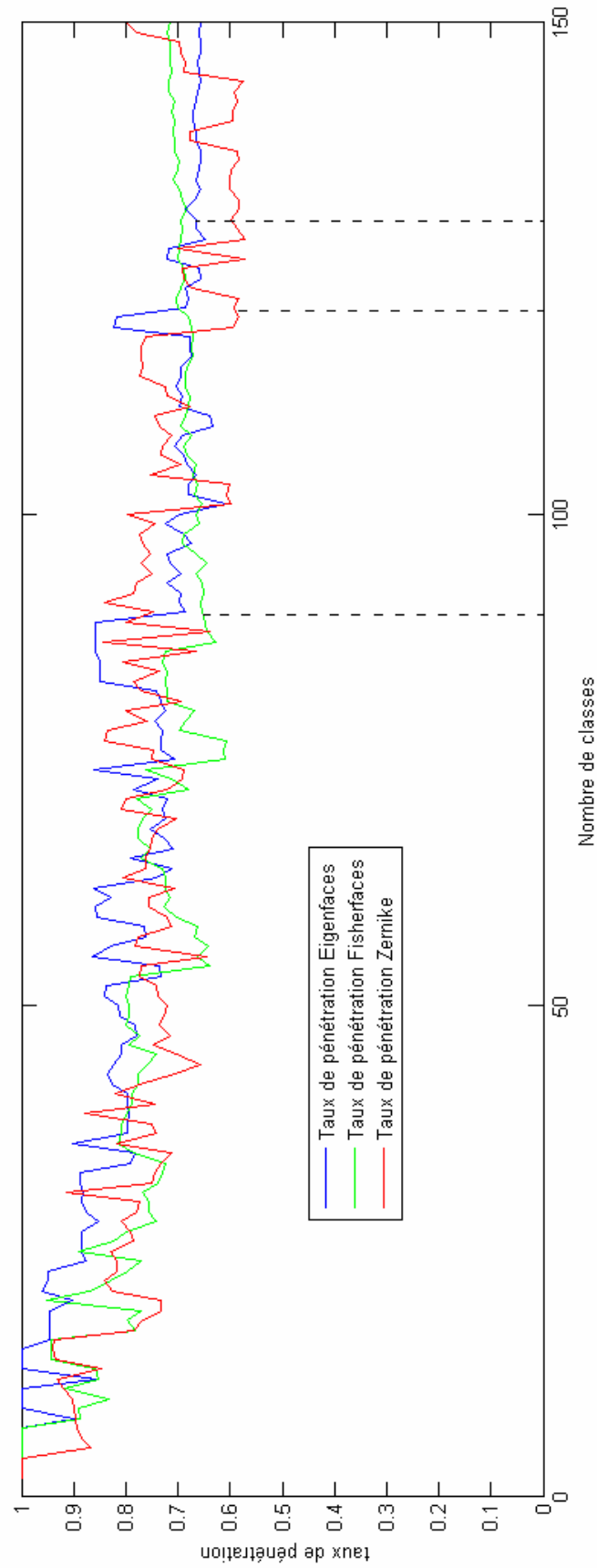


Figure 6.7 : Taux de pénétration des *Eigenfaces* (bleu), des *Fisherfaces* (vert) et des moments de Zernike (rouge) suite à une classification par les K-moyennes.

Nous remarquons un taux de pénétration comparable entre les trois méthodes de représentation. Mais, les moments de Zernike présentent un meilleur comportement pour un nombre de classes compris entre 120 et 140 et minimisent le taux de pénétration R à moins de 60%. En effet, pour parvenir à une réduction assez conséquente de la galerie, il s'agit de choisir le nombre de classe qui minimise le taux de pénétration. Cependant, pour assurer la robustesse et la portabilité des résultats sur d'autres données de test, nous recommandons le choix d'un nombre de classes qui assure une certaine stabilité du taux de pénétration en ajoutant ou en diminuant une classe. D'autres techniques plus poussées peuvent être utilisées pour valider un nombre de classes donné. Gordon [GC88] et Milligan [MC96] ont proposé une technique qui mesure la stabilité des résultats obtenus d'une classification par rapport aux différentes perturbations que les données peuvent subir. Ces perturbations consistent à retirer un élément de l'ensemble des données et mesurer l'influence du retrait de cet élément sur la classification. La stabilité de la classification est mesurée par l'écart entre la structure initiale et la structure obtenue sur les données bruitées ou par la variation d'un critère mesuré sur ces deux structures.

Ainsi nous retenons un nombre de classes égal à 120 pour les moments de Zernike ce qui donne un taux de pénétration de l'ordre de 59%. Pour les *Eigenfaces*, nous retenons un nombre de 130 classes ce qui correspond à un taux de pénétration de l'ordre de 66%. Finalement, pour les *Fisherfaces*, nous retenons un nombre autour de 90 classes impliquant un taux de pénétration de l'ordre de 65%.

6.5.3.2 Réduction de la galerie par l'algorithme des centres mobiles de Forgy

Les résultats obtenus dans ce paragraphe se basent sur l'utilisation de l'algorithme des centres mobiles de Forgy pour classifier les attributs faciaux extraits par les *Eigenfaces*, les *Fisherfaces* et par les moments de Zernike.

En analysant la figure 6.8, nous remarquons que le taux de pénétration est optimum pour un nombre de classes K autour de 100 à la fois pour les attributs *Eigenfaces* et *Fisherfaces*. Les taux de pénétration sont de l'ordre de 70% ce qui donne une simplification de plus de 30% de la galerie. Quant aux moments de Zernike (courbe rouge), la simplification de la base par les centres mobiles de Forgy est beaucoup moins intéressante que celle effectuée avec l'algorithme des K-moyennes. Pour un nombre de classe autour de 100, le taux de pénétration atteint seulement les 80% ce qui permet de simplifier uniquement un cinquième de la galerie. En effet, les vecteurs attributs formés par les moments de Zernike sont d'une taille de 128 composantes et constituent une dimensionnalité assez conséquente pour le problème de classification.

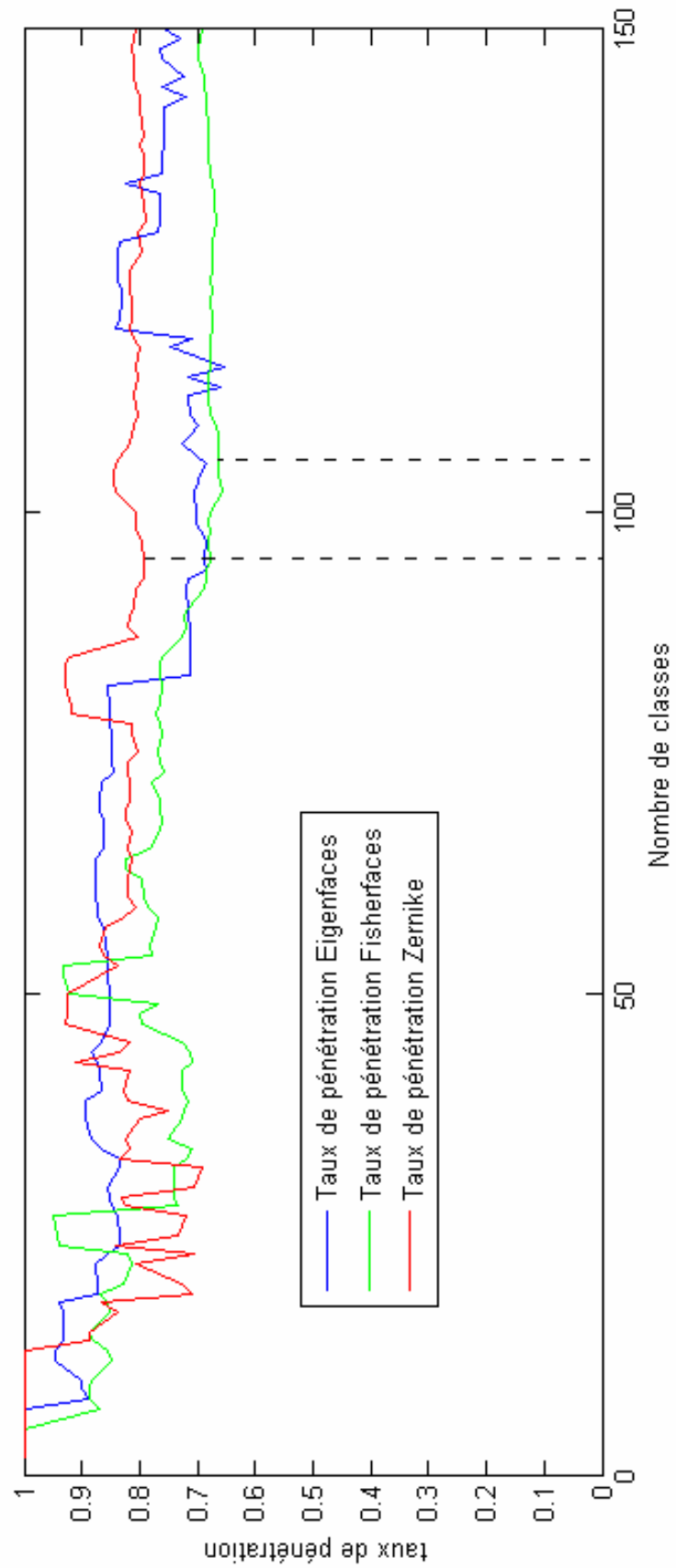


Figure 6.8 : Taux de pénétration des *Eigenfaces* (bleu), des *Fisherfaces* (vert) et des moments de Zernike (rouge) suite à une classification automatique par les centres mobiles de Forgy.

6.5.3.3 Réduction de la galerie par la classification BIRCH

Pour produire un nombre de classes souhaité à l'aide de l'algorithme BIRCH, il faut ajuster à la fois les paramètres B et L relatifs respectivement à la taille d'un nœud non feuille et à la taille d'un nœud feuille ainsi que le seuil d'absorption T . Il est à noter que seuls les nœuds feuilles résument effectivement les points de données et constituent les classes d'individus. En se basant sur les classifications effectuées avec les K-moyennes et les centres mobiles de Forgy, le nombre optimal de classes se situe entre 50 à 120 classes pour les différents attributs utilisés. Ceci est le seul a priori dont nous disposons pour ajuster au mieux les paramètres de BIRCH. Etant donné le nombre de personnes de la galerie $N = 1880$, une classe peut contenir en moyenne de 15 à 38 individus. Ainsi, nous avons fait varier, initialement, le facteur L entre ces valeurs de 5 à 20 unités avec un seuil d'absorption T nul et un facteur bifurquant B égal à 20. Le tableau 6.4 illustre le nombre de classes (K), le nombre de classes retenues (P) et le taux de pénétration (R) correspondant au nombre moyen d'identités retenues de la galerie à partir de l'ensemble des images de test.

	L	20	10	9	8	7	6	5
<i>Eigenfaces</i>	K	5	26	34	47	50	74	114
	P	4	14	24	29	48	46	83
	R	0,89	0,71	0,81	0,74	0,96	0,74	0,76
<i>Fisherfaces</i>	K	5	23	27	32	52	69	109
	P	3	21	21	20	27	45	79
	R	0,90	0,98	0,93	0,82	0,76	0,82	0,83
Moments de Zernike	K	14	45	58	59	106	139	185
	P	11	34	38	45	72	120	123
	R	0,93	0,86	0,82	0,90	0,81	0,93	0,78

Tableau 6.4 : Résultats de réduction de la galerie par BIRCH avec un seuil d'absorption $T=0$.

Il est à noter que la réorganisation des points de données autour des centroïdes les plus proches laisse certaines classes initiales vides et sans représentant effectif. Ainsi, cette étape a tendance à faire diminuer le nombre des classes générées par l'arbre CF .

Nous remarquons un comportement non monotone des taux de pénétration R en fonction de la taille des nœuds feuilles L . Ce comportement est difficilement interprétable. Cependant, il est possible d'avancer l'hypothèse que les microclasses générées par l'arbre- CF sont très sensibles à la taille des nœuds en opérant avec un seuil d'absorption $T = 0$.

Nous retenons, parmi les meilleures performances du tableau 6.4 et pour chaque type d'attributs, les résultats suivants:

- *Eigenfaces* : pour $L = 10$; $R = 0,71$
- *Fisherfaces* : pour $L = 7$; $R = 0,76$
- Moments de Zernike : pour $L = 5$; $R = 0,78$

Dans un deuxième temps, nous avons fait augmenter le seuil d'absorption T pour générer des classes plus compactes et mieux représentatives de la distribution des données. Nous avons ajusté pour cela les facteurs bifurquant B et L à 5 unités. Proportionnellement à la taille et à l'ordre de grandeur des vecteurs attributs utilisés, nous avons essayé quelques seuils d'absorption T . Le tableau 6.5 présente les différents seuils retenus pour chaque type d'attributs ainsi que les taux de pénétration relatifs.

attributs	seuil d'absorption T	K	P	R
<i>Eigenfaces</i>	1,5	89	5	71%
<i>Fisherfaces</i>	0,5	82	41	62,1%
moments de Zernike	26000	110	55	67,65%

Tableau 6.5 : Résultats de réduction de la galerie par BIRCH avec un seuil d'absorption $T > 0$.

6.5.3.4 Comparaison des différentes méthodes de partitionnement

Les vecteurs attributs de dimension g extraits par chacune des méthodes de description des images de visage sont assimilés à des points de données multidimensionnels. Ces points de données constituent l'entrée du module de classification non supervisée.

En nous basant sur le critère du taux de pénétration, nous avons choisi, pour chaque méthode de classification non supervisée et pour chaque type de description des images faciales, le nombre de classe optimal qui minimise ce taux. Les tableaux 6.6 et 6.7 décrivent le nombre de classes retenues et le taux de pénétration R des différentes stratégies de classification.

descripteurs \ classifieurs	K-moyennes	Centres mobiles	BIRCH
<i>Eigenfaces</i>	130	95	89
<i>Fisherfaces</i>	90	103	82
Moments de Zernike	120	95	110

Tableau 6.6 : Nombre de classes K pour les différentes stratégies de classification.

Attributs \ Classifieurs	K-moyennes	Centres mobiles de Forgy	BIRCH
<i>Eigenfaces</i>	66,55%	68,96%	71%
<i>Fisherfaces</i>	65,37%	66%	62,1%
Moments de Zernike	58,38%	79,24%	67,65%

Tableau 6.7 : Taux de pénétration R des différents classifieurs.

La classification par K-moyennes élaborée sur les moments de Zernike semble être la meilleure stratégie pour partitionner la galerie des images de visages. En effet, cette stratégie assure un taux de pénétration de l'ordre de 58% et par conséquent une réduction conséquente

de la galerie. En contre partie, les moments de Zernike utilisés avec les centres mobiles de Forgy fournissent des résultats très moyens avec une réduction de seulement 79% de la galerie.

Le taux de pénétration des K-moyennes et des centres mobiles de Forgy en utilisant les *Eigenfaces* ainsi que les *Fisherfaces* sont comparables aux alentours de 66%. Dans ces quatre cas, la réduction d'un tiers de la galerie est assurée. Cependant la méthode BIRCH de classification, présente des résultats instables à l'image de ceux fournis par le tableau 6.4. Certes, nous avons déjà avec seulement quelques essais, un meilleur comportement de cette méthode appliquée sur les vecteurs attributs *Fisherfaces* (62% de taux de pénétration) qu'en utilisant les K-moyennes et les centres mobiles de Forgy.

Un ajustement plus judicieux des paramètres N et L ainsi que du seuil d'absorption T permettra une meilleure exploitation de la méthode BIRCH de classification non supervisée. L'ajustement de ces paramètres peut être mieux appréhendé pour générer des classes plus compactes et mieux séparées. Cela nécessite, sans doute, une batterie de tests pour ajuster surtout le seuil d'absorption T en fonction des données d'apprentissage et de validation.

Ceci mettra en cause cette technique pour le développement d'un système automatique de reconnaissance de visage. Un tel système nécessite des méthodes suffisamment stables et universelles, c'est-à-dire qui ne dépendent pas de l'ajustement d'un nombre important de paramètres en fonction des caractéristiques des bases de visages utilisées.

6.6 Etude d'indépendance des différents classifieurs

Nous développons essentiellement, dans cette partie, les tableaux de contingence pour mesurer le degré de corrélation de deux partitions résultantes de deux classifieurs différents. Le tableau de contingence, appelé aussi matrice de confusion dans la terminologie de l'apprentissage supervisé, est un outil servant à mesurer la qualité d'une classification. Chaque colonne de la matrice représente le nombre d'occurrences d'une classe estimée, tandis que chaque ligne représente le nombre d'occurrences d'une classe réelle. En normalisant cette matrice, une classification est d'autant meilleure que sa matrice de confusion s'approche d'une matrice diagonale.

Nous avons développé cet outil statistique pour mesurer l'indépendance deux à deux des classifications non supervisées mises en œuvre. Nous procédons dans l'expérimentation suivante à la partition de la galerie en 10 classes. La matrice générée (Tableau 6.8) illustre la dépendance entre les classes des vecteurs attributs *Eigenfaces* en lignes et ceux des vecteurs *Fisherfaces* en colonnes. Il est à remarquer, par exemple, une forte dépendance entre la première classe issue de la partition *Eigenfaces* et celle de la partition *Fisherfaces* avec 224 individus en commun. D'autres classes d'individus générées par une des deux méthodes de partition sont plus ou moins répartis entre les classes de l'autre partition. Les individus de la classe numéro 4 de la partition *Eigenfaces*, par exemple, sont relativement répartis entre plusieurs classes de la partition *Fisherfaces*, alors que la classe numéro 8 de la première partition est exclusivement contenu dans la classe numéro 2 de la deuxième partition.

Ce tableau de données associe deux descripteurs qualitatifs de deux classifications différentes. Une case de la $i^{\text{ème}}$ ligne et de la $j^{\text{ème}}$ colonne représente le nombre d'individus appartenant à la fois à la classe i de la partition *Eigenfaces* et à la classe j de la partition *Fisherfaces*. Les totaux des lignes x et des colonnes y sont appelés "totaux marginaux" et indiquent la taille ou

la fréquence de chaque classe indépendamment des autres. N étant le nombre total des individus de la galerie qui est égal à 1880 personnes.

		<i>Fisherfaces</i>										x
		1	2	3	4	5	6	7	8	9	10	
<i>Eigenfaces</i>	1	224	0	0	0	0	0	0	0	13	0	237
	2	0	71	0	0	0	0	7	64	0	0	142
	3	0	0	0	1	1	48	0	0	0	72	122
	4	0	0	153	59	5	0	1	8	0	3	229
	5	0	0	0	13	13	0	0	0	133	0	159
	6	40	0	0	0	17	1	0	0	180	21	259
	7	0	4	9	0	2	0	70	7	0	0	92
	8	0	45	0	0	0	0	0	0	0	0	45
	9	0	19	31	1	0	0	2	134	0	0	187
	10	0	0	8	113	221	0	2	0	43	21	408
y	264	139	201	187	259	49	82	213	369	117	1880	

Tableau 6.8 : Tableau de contingence de deux partitions différentes de la galerie.

En analyse de données le test de χ^2 permet de mesurer l'indépendance de deux variables qualitatives différentes. Ce test consiste à mesurer l'écart entre une situation observée et une situation théorique pour en déduire le degré ou le risque de dépendance (corrélation) des deux variables mesurées. Dans notre cas, la valeur observée " O " correspond au tableau de contingence établi alors que la situation théorique se résume au calcul des espérances " E ". Dans chaque case du tableau, l'espérance E_{ij} est égale à la somme des éléments de la ligne i multipliée par la somme des éléments de la colonne j , divisée par N , soit $(x_i \cdot y_j)/N$. Le tableau des espérances E peut être interprété comme les valeurs attendues en cas d'indépendance des deux classifications. Le test de χ^2 mesure ainsi l'indépendance des deux classifications effectuées comme suit :

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (6.7)$$

6.6.1 Indépendance des attributs

Les vecteurs attributs conçus pour la classification des identités sont extraits pour chaque type de description développée des images faciales, à savoir les *Eigenfaces*, les *Fisherfaces*, et les moments de Zernike. Les paramètres de la classification et notamment le nombre de classes qui optimise le taux de pénétration pour chaque type d'attribut et chaque méthode de classification sont développés dans le reste des expériences. Nous avons calculé le test de χ^2 pour les trois configurations des vecteurs attributs suivantes : *Eigenfaces* vs *Fisherfaces*, *Eigenfaces* vs moments de Zernike, et *Fisherfaces* vs moments de Zernike. La classification automatique est effectuée avec l'algorithme des K-moyennes. Les résultats sont présentés dans le tableau 6.9.

Combinaison d'attributs	PCA vs LDA	PCA vs Zernike	LDA vs Zernike
Test de χ^2	65458	105526	53422

Tableau 6.9 : Test d'indépendance de χ^2 entre les partitions de différentes méthodes de représentation.

En analyse de données, où l'objectif est de réduire la dimensionnalité du problème, on cherche les classes dépendantes, c'est-à-dire les classes ayant un risque d'indépendance minimal. Dans ce cas, les valeurs du test de χ^2 sont comparées avec celles d'un tableau de référence pour mesurer le risque d'indépendance de deux variables mesurées. Cependant, dans notre cas d'exercice, nous comparons les valeurs du test entre elles afin de faire ressortir les configurations les plus dé-corrélées. En effet, nous nous intéresserons plus tard à la fusion de classifieurs différents qui soient le plus dé-corrélés l'un de l'autre pour pouvoir améliorer les résultats de réduction de la galerie. Ainsi, les valeurs minimales du test de χ^2 indiquent une plus grande indépendance entre les partitions correspondantes.

Il est clair, de ce fait, que les partitions obtenues par les *Fisherfaces* et les moments de Zernike sont les plus dé-corrélées parmi les trois configurations évaluées. Les partitions des *Eigenfaces* et des *Fisherfaces* sont un peu plus dépendantes, alors que celles des *Eigenfaces* et des moments de Zernike sont beaucoup plus dépendantes.

6.6.2 Indépendance des méthodes de classification et des mesures de similarité

Nous avons réalisé la même expérimentation précédente, tout d'abord, entre les partitions obtenues avec les différentes méthodes de classification développées, puis, entre les mesures de similarités L_1 , L_2 et Cosinus en opérant une classification avec l'algorithme des K-moyennes et les vecteurs attributs *Eigenfaces*. Le tableau 6.10 présente les résultats de dépendance entre les partitions : K-moyennes vs centres mobiles de Forgy, K-moyennes vs BIRCH et centres mobiles de Forgy vs BIRCH. Le tableau 6.11 présente, à son tour, les résultats de dépendance entre les partitions réalisées avec les K-moyennes à travers les mesures de similarité suivantes : distance Euclidienne (L_2) vs distance de Manhattan (L_1), distance Euclidienne vs distance Cosinus et distance de Manhattan vs distance Cosinus.

Combinaison d'attributs	K-moyennes vs Centres mobiles de Forgy	K-moyennes vs BIRCH	Centres mobiles de Forgy vs BIRCH
Test de χ^2	114402	44362	42109

Tableau 6.10 : Test d'indépendance de χ^2 entre les partitions de différentes méthodes de classification.

Combinaison d'attributs	L_2 vs L_1	L_2 vs Cosinus	L_1 vs Cosinus
Test de χ^2	123775	82963	73559

Tableau 6.11 : Test d'indépendance de χ^2 entre les partitions de différentes mesures de similarités dans la classification des attributs *Eigenfaces*.

Les partitions obtenues par les méthodes des K-moyennes et celle des centres mobiles de Forgy semblent être les plus dépendantes avec une valeur du test de χ^2 de 114402. D'autre part, la partition générée avec l'algorithme BIRCH est beaucoup moins dépendante de celle des K-moyennes ainsi que celle des centres mobiles de Forgy. Ainsi une combinaison entre ces deux dernières paires de classifieurs semble être plus probante et fructueuse en termes de réduction de la galerie. Quant aux mesures de similarité employées pour effectuer la partition de la galerie, les distances de Manhattan et Cosinus semblent être les plus indépendantes parmi les configurations évaluées. D'autre part la distance de Manhattan et la distance Euclidienne fournissent les partitions les plus corrélées vis-à-vis des autres combinaisons étudiées (cf Tableau 6.11).

6.6.3 Choix de la distance de similarité

Bien que la distance de Manhattan donne souvent de meilleurs résultats que la distance Euclidienne dans un processus d'identification, nous avons remarqué un meilleur comportement de la distance Euclidienne quant aux rangs d'identification totale (*TMR*) ainsi qu'en mesurant les taux de pénétration. En outre, sur la base de ces deux derniers indices, nous avons remarqué un meilleur comportement de la distance Euclidienne par rapport à la distance Cosinus.

Dans la communauté de la reconnaissance des formes, le choix d'une distance de similarité est souvent expliqué et argumenté par rapport à l'espace d'attributs et aux points de données utilisés. Nous avons remarqué entre autre, avec les mêmes vecteurs attributs *Eigenfaces*, que la distance de Manhattan fournit de meilleurs résultats d'identification des individus avec 85,25% de taux d'identification contre 79% avec la distance Euclidienne. Cependant la distance Euclidienne dépasse cette dernière dans le processus d'identification des classes (*C-Identification*). Ces performances nous ont poussés à utiliser la distance Euclidienne dans ce dernier processus. Peut on avancer, de ce fait, que la distance Euclidienne est mieux adaptée à un processus global tel que la reconnaissance de classes d'objets qu'à un processus spécifique d'identification d'individus. Les résultats obtenus plaident en tout cas pour cette constatation.

6.7 Agrégation et combinaison de classifieurs

Maltoni et al. [MMJP03] ont catégorisé différents schémas de combinaison de classifieurs basés sur l'architecture d'intégration, le niveau de fusion et les stratégies de fusion des classifieurs individuels. Les auteurs montrent que les stratégies de fusion peuvent dégrader la précision et les performances de la classification. La combinaison des classifieurs hautement corrélés peut induire une baisse des résultats du fait qu'il n'y a pas une information supplémentaire et les erreurs peuvent s'accumuler. Pour assurer qu'une combinaison améliore les performances, Prabhakar et Jain [PJ02] ont proposé une sélection automatique de classifieurs pour éliminer initialement les classifieurs hautement corrélés. D'autre part, la méthode de combinaison ou de fusion des sorties des classifieurs joue un rôle important dans les performances du système : une combinaison mal conçue peut dégrader les performances du système de même que la combinaison de classifieurs redondants. En outre, Prabhakar et Jain [PJ02] argumentent, dans le contexte de classification des empreintes digitales, que la combinaison de deux classifieurs moyens est susceptible d'améliorer largement les résultats alors que celle de deux bons classifieurs peut induire une toute légère amélioration. Ainsi les deux meilleurs classifieurs individuels ne forment pas nécessairement la meilleure paire de classifieur à retenir. Dans cette perspective, la méthode de classification Adaboost qui

effectue par itérations successives une combinaison de faibles classifieurs apparait comme une intéressante alternative dans le cadre d'un apprentissage supervisé [FS95].

6.7.1 Les architectures de combinaison

Les principales architectures de combinaison de classifieurs sont les architectures en parallèle, en série et hiérarchique (voir figure 6.9) [MMJP03]. Dans une architecture parallèle, tous les classifieurs fonctionnent indépendamment l'un de l'autre et les résultats des classifieurs individuels sont combinés.

Dans un schéma en cascade, les classifieurs individuels opèrent l'un après l'autre dans une séquence. Le nombre de classes possibles dans lesquelles une forme donnée peut être assignée est successivement réduit d'autant que les classifieurs dans la séquence sont appelés. Par exemple, dans une procédure d'identification impliquant une base de données de N classes, le premier classifieur peut réduire le nombre des classes potentielles à une fraction de N ; le second classifieur peut encore réduire ce nombre de classes possible à un nombre plus petit et ainsi de suite. Typiquement, les classifieurs les moins coûteux au niveau calculatoire et d'extraction des caractéristiques et qui ont relativement une faible précision sont employés en premier et sont suivi par les classifieurs les plus précis et les plus coûteux.

Dans le schéma hiérarchique, les classifieurs individuels sont combinés suivant une structure arborescente.

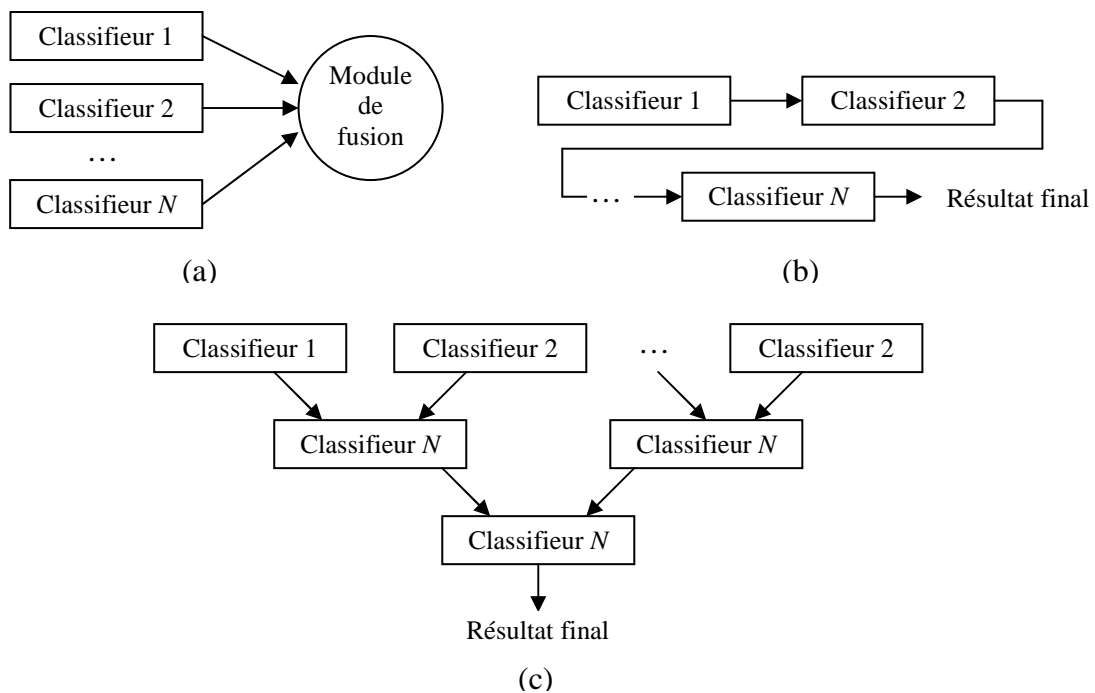


Figure 6.9 : Combinaison a) en parallèle, b) en série, c) hiérarchique de N classifieurs

6.7.2 Les différents niveaux de fusion pour un système multimodal

En généralisant le processus d'identification et de reconnaissance à plusieurs modalités, la fusion des données extraites peut se faire sur différents niveaux [RJ03]. La combinaison de l'information peut intervenir (a) au niveau des échantillons saisis par les capteurs, (b) au niveau de l'extraction des caractéristiques, (c) au niveau du score de l'appariement et (d) au niveau de la prise de décision. Notons que la fusion aux niveaux (a) et (b) se produit avant

que la procédure d'appariement soit mise en œuvre, alors que les niveaux (c) et (d) interviennent après l'appariement des modèles. Bien que l'intégration soit possible à ces différents niveaux, la fusion au niveau du score et de la décision est une des méthodes le plus généralement utilisée.

Les différents niveaux de la fusion pour un système multimodal incluent:

- **Niveau de décision :** Chaque processus biométrique produit son propre résultat booléen. Le processus de fusion les assemble par un algorithme de combinaison comme ET, OU, etc.
- **Niveau du score :** Chaque processus biométrique produit typiquement un score d'appariement ou plusieurs scores. Le processus de fusion combine ces derniers dans un résultat unique, qui est alors comparé au seuil d'acceptabilité du système. Si nous procédons à l'approche de classification pour la fusion des scores, la sortie est alors une décision directe et non des scores.
- **Niveau des caractéristiques :** Chaque processus biométrique produit une série de caractéristiques. Le processus de fusion combine ces collections de caractéristiques en un unique ensemble ou vecteur de caractéristiques.
- **Niveau des échantillons :** Chaque processus biométrique produit une collection d'échantillons à partir des différents capteurs. Le processus de fusion combine ces données dans un simple échantillon.

La fusion biométrique combine plusieurs échantillons ou caractéristiques biométriques acquises par une seule ou différentes modalités ou technologies. Le but est d'identifier ou authentifier des individus plus efficacement qu'en utilisant un seul échantillon ou modalité. Dans un processus de reconnaissance multimodal, plus tôt la combinaison est appliquée plus elle est efficace [JNR05]. Par exemple, une intégration au niveau des caractéristiques fournit de meilleurs résultats qu'au niveau des scores de confiance ou au niveau de la décision. Ceci est expliqué par le fait que la représentation des caractéristiques donne la plus riche information, suivi par les scores de confiance d'un classifieur, et les résultats de décision contiennent la plus petite quantité d'information pour la décision à prendre. Cependant il est plus difficile de combiner les classifieurs au niveau des caractéristiques parce que les espaces de caractéristiques des différents classifieurs sont inconnus et la représentation des caractéristiques peut ne pas être compatible. D'autre part, l'aspect confidentiel et propriétaire des valeurs dans l'espace des caractéristiques a freiné la recherche quant à la combinaison au niveau des caractéristiques. Dans de tels cas, les niveaux des scores et de décision sont les seules alternatives.

6.7.3 Notre stratégie de fusion

Nous proposons ici une approche d'agrégation et de fusion de deux classifieurs afin de réduire sensiblement la taille de la galerie pour une identification ultérieure. Nous optons pour une architecture de combinaison en parallèle. Le coût global de l'opération en termes de complexité calculatoire s'élèvera au coût du classifieur le plus gourmand en temps de calcul ajouté au coût de l'étape de fusion. Chaque classifieur propose une partition de la galerie suivant laquelle un ensemble d'identités sera retenu et l'ensemble complémentaire rejeté de la galerie. La fusion des classifieurs intervient, ainsi, au niveau de la décision et utilise l'information « retenue / rejetée » produite par chaque classifieur sur chaque identité de la galerie afin de générer la décision finale.

Chaque classifieur est conçu de façon à produire une simplification aussi fiable et sûre dans le sens où il retient dans la galerie, à tous les coups, l'identité recherchée par le système. Ceci nous permet d'adopter l'algorithme de combinaison du ET logique pour décider des identités retenues et des identités rejetées. Une identité n'est retenue par le système de fusion que si elle l'est par les deux classifieurs mis en jeux. Cette règle de fusion implique éventuellement une diminution du nombre de personnes retenues dans la galerie pour la suite de la recherche. Ceci entraîne un plus petit taux de pénétration et une plus grande simplification que ceux de chacun des classifieurs. Par contre, bien que nous essayions de l'annuler pour chaque classifieur, le risque de rejet d'une personne recherchée a tendance à s'accroître.

L'exemple développé dans le tableau 6.12 présente la matrice de confusion d'un classifieur basé sur les attributs *Eigenfaces* (PCA) et d'un classifieur basé sur les attributs *Fisherfaces* (LDA). La cellule en gras indique le nombre moyen de personnes retenues par notre module de fusion de ces deux classifieurs sur une base de 3111 images de test. Alors que les deux classifieurs basés sur les attributs *Eigenfaces* et les attributs *Fisherfaces* permettent de retenir respectivement 1251 et 1229 identités parmi les 1880 identités de la galerie, la fusion de ces deux classifieurs réduit le nombre moyen des personnes retenues à 1122,8.

		LDA		Total
		retenues	rejetées	
PCA	retenues	1122,8	128,46	1251,26
	rejetées	106,2	522,54	628,74
Total		1229	651	1880

Tableau 6.12 : Matrice de confusion entre le classifieur basé sur les attributs *Eigenfaces* (en lignes) et le classifieur basé sur les attributs *Fisherfaces* (en colonnes).

6.7.4 Résultats de la fusion

Sur la base des études effectuées dans la section précédente portant sur l'indépendance des partitions, nous avons repris les mêmes paires de classifieurs du tableau 6.9 élaborées avec les attributs des *Eigenfaces*, des *Fisherfaces* et les moments de Zernike. Le tableau 6.13 présente le nombre de personnes retenues ainsi que le taux de pénétration des classifieurs fusionnés.

Fusion des classifieurs	PCA vs LDA	PCA vs Zernike	LDA vs Zernike
Nombre d'identités retenues	1122,8	1018,2	993,28
Taux de pénétration	59,7%	54,16%	52,8%

Tableau 6.13 : Nombre de personnes retenues entre les différentes paires de classifieurs.

Nous rappelons que tous les classifieurs sont développés avec des attributs différents mais avec la même méthode de partitionnement : les K-moyennes. Nous rappelons que les deux classifieurs qui génèrent les partitions les plus décorréliées parmi les paires testées sont ceux basés sur les attributs *Fisherfaces* et les moments de Zernike. Ces deux classifieurs, améliorent considérablement le taux de pénétration jusqu'à 52,8%. Cependant, la paire de classifieurs basés sur les attributs *Eigenfaces* et les moments de Zernike, qui génèrent les

partitions les plus dépendantes donnent de meilleurs résultats que ceux des classifieurs basés sur les attributs *Eigenfaces* et *Fisherfaces*. En effet, comme le montre la figure 6.8, le partitionnement par les K-moyennes des moments de Zernike fournit de meilleurs taux de pénétration que ceux obtenus avec les *Eigenfaces* et les *Fisherfaces*. Les moments de Zernike réduisent à eux seuls plus de 40% de la galerie avec un nombre d'identités retenues moyen de 1098 sur les 1880 personnes préenregistrées. Mais, nous remarquons que le gain de réduction en employant la fusion des *Eigenfaces* et des moments de Zernike est bien minimal par rapport aux moments de Zernike seul. Ceci explique, sans doute, la corrélation des partitions de ces deux descripteurs.

Nous pouvons conclure en dehors des performances intrinsèques de chaque classifieur, que plus les classifieurs sont dé-corrélés, plus leur fusion implique une plus grande réduction de la galerie. Dans cette perspective, nous avons combiné deux classifieurs encore plus dé-corrélés, non seulement à travers les attributs mais aussi par les méthodes de classification développées. Le premier classifieur effectue la classification des vecteurs attributs *Fisherfaces* par la méthode BIRCH, alors que le deuxième classifieur les moments de Zernike par l'algorithme des K-moyennes. Le tableau 6.14 présente la matrice de confusion des partitions retenues et rejetées de ces deux derniers classifieurs.

		Zernike K-moyennes		Total
		retenues	rejetées	
LDA BIRCH	retenues	949,03	218,43	1167,46
	rejetées	148,58	563,96	712,54
Total		1097,61	782,39	1880

Tableau 6.14 : Matrice de confusion entre le classifieur basé sur les attributs *Fisherfaces* et la classification BIRCH (en lignes) et le classifieur basé sur les moments de Zernike et le partitionnement par les K-moyennes (en colonnes).

Nous remarquons que cette combinaison simplifie encore plus les identités de la base et engendre un espace de recherche encore plus réduit de 949 identités. Ceci revient, à peu près, à 50% de taux de pénétration ce qui correspond à la réduction de la moitié de la taille de la galerie sans risquer de perdre l'identité recherchée.

6.8 Identification finale

Nous montrons dans cette dernière section les performances enregistrées par notre approche d'identification en termes de précision et de vitesse d'exécution. La comparaison des résultats obtenus avec ceux employant les mêmes techniques fournies par la méthode classique d'identification donne une évaluation objective de notre système d'identification.

A la suite de l'étape de partitionnement de la galerie que nous avons effectuée avec une classification non supervisée, le déroulement de l'identification des individus se fait en deux phases. La première phase se rapporte à la recherche des classes les plus proches de l'identité requête. Il s'agit d'organiser ces classes puis de sélectionner l'ensemble des identités contenues dans les P classes retenues. Une fois l'ensemble des identités retenues pour la suite de la recherche sont sélectionnées, la deuxième phase intervient et consiste à opérer

l'identification des identités requêtes. La figure 6.10 illustre ces deux phases de recherche en ligne dans notre processus d'identification.

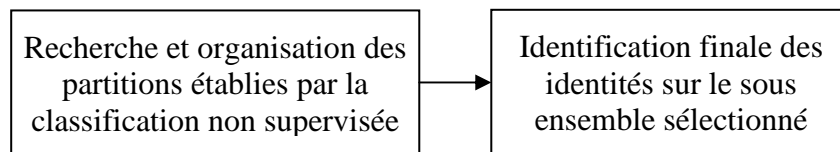


Figure 6.10 : Déroulement de l'étape d'identification (En ligne) en aval de la classification non supervisée.

6.8.1 Précision de l'identification

Plusieurs stratégies de combinaisons possibles peuvent être élaborées entre le processus de classification non supervisée et les deux phases d'identification illustrées par la figure 6.9. En effet, l'utilisation de méthodes de classification non supervisée distinctes ainsi que différents descripteurs et mesures de similarités pour chacune de ces étapes de classification et de recherche offre une large panoplie de possibilités.

Nous avons exclu dès le début l'utilisation d'une même méthode de description, comme les *Eigenfaces*, pour établir à la fois les vecteurs attributs de classification et ceux d'identification. Bien que ces deux vecteurs extraits soient distincts, ils présentent tout de même une forte corrélation. De ce fait, les identités qui engendrent des faux appariements lors de l'identification finale se trouvent par conséquent toutes retenues par la phase de recherche des partitions. L'amélioration de la précision de l'identification est très marginale dans ce cas et sera compromise par le taux d'erreur de la recherche *RER*.

Nous avons comparé les taux d'identification de l'approche classique (identification directe sur toute la galerie) avec ceux de notre approche pour la majorité des stratégies restantes. Aucune de ces stratégies n'arrive à améliorer le taux d'identification. La seule configuration permettant d'améliorer très légèrement le taux d'identification (de 81,67% à 81,71%) est celle qui établit la classification des vecteurs attributs *Fisherfaces* avec les K-moyennes. L'identification finale est réalisée avec les vecteurs attributs *Eigenfaces* dédiés. Cependant, notre approche assure au pire des cas les performances de l'identification classique. Pour examiner le comportement des réponses à un rang donné de notre système d'identification, nous avons tracé la courbe de scores cumulés. Les figures 6.11 6.12 et 6.13 illustrent les courbes de scores cumulés de trois combinaisons différentes. Sur chaque figure, nous traçons en rouge la courbe de scores cumulés de notre approche d'identification et en bleu celle de l'identification classique. Nous avons combiné les différentes méthodes pour la classification, l'identification des classes et pour l'identification finale des individus. Etant donné que les meilleurs résultats d'identification sont fournis par les *Fisherfaces*, la meilleure combinaison illustrée en rouge par la figure 6.12 est réalisée par la combinaison des *Eigenfaces* pour la classification et les *Fisherfaces* pour l'identification.

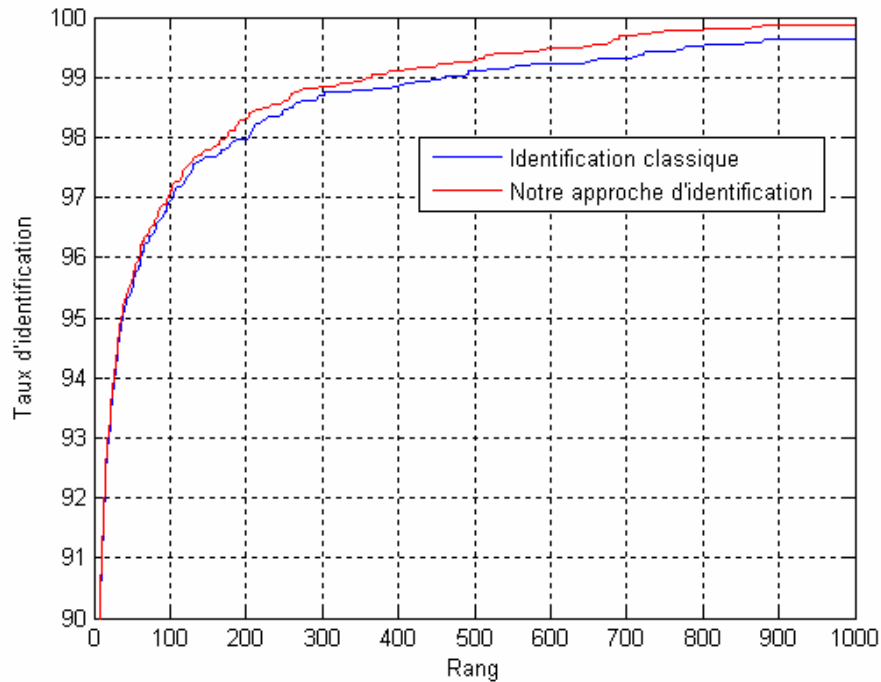


Figure 6.11 : Comparaison des courbes de scores cumulés de notre approche et celle de l'identification classique. La classification est réalisée avec BIRCH sur les attributs *Fisherfaces* ($g = 4$) alors que l'identification est réalisée avec les attributs *Eigenfaces* ($g = 131$).

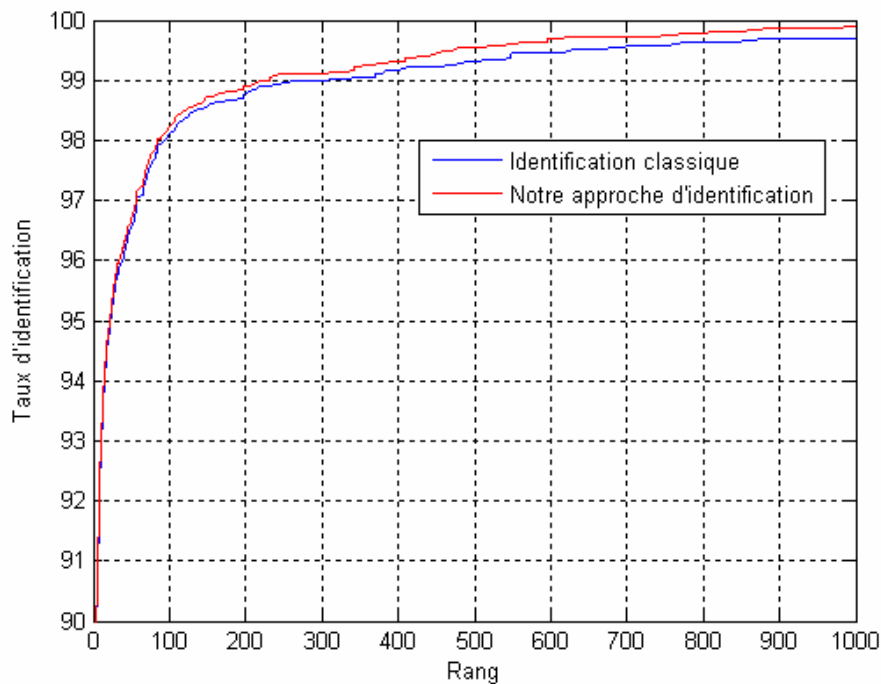


Figure 6.12 : Comparaison des courbes de scores cumulés de notre approche et celle de l'identification classique. La classification est réalisée avec les K-moyennes sur les attributs *Eigenfaces* ($g = 16$) alors que l'identification est réalisée avec les *Fisherfaces* ($g = 67$).

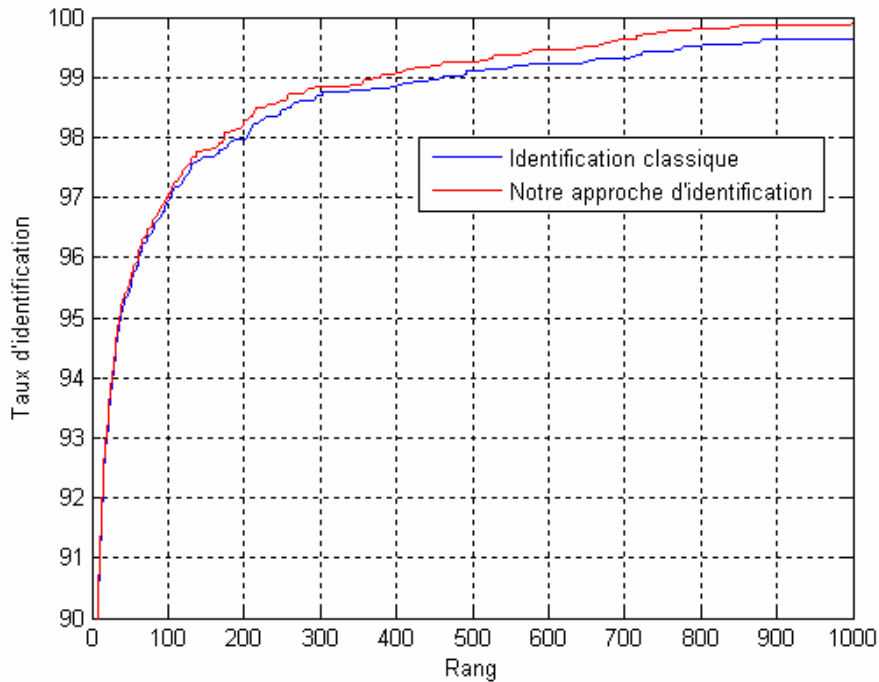


Figure 6.13 : Comparaison des courbes de scores cumulés de notre approche et celle de l'identification classique. La classification est réalisée avec les K-moyennes sur les moments de Zernike (ordre = 14) alors que l'identification est réalisée avec les *Eigenfaces* ($g = 131$).

La meilleure amélioration des résultats par rapport à l'identification classique se traduit par la courbe de scores cumulés la plus décalée au dessus de celle de l'identification classique. En se basant sur l'étude de l'indépendance des partitions établies précédemment, nous avons fusionné les deux partitions explicitées dans le tableau 6.14. La première partition est établie avec la méthode BIRCH sur les attributs *Fisherfaces*. La deuxième partition est réalisée par la classification K-moyennes des moments de Zernike. L'identification par les *Eigenfaces* du sous-ensemble d'identités retenu par la fusion de ces deux partitions est comparée à l'identification classique par les mêmes vecteurs attributs *Eigenfaces* (cf Figure 6.14).

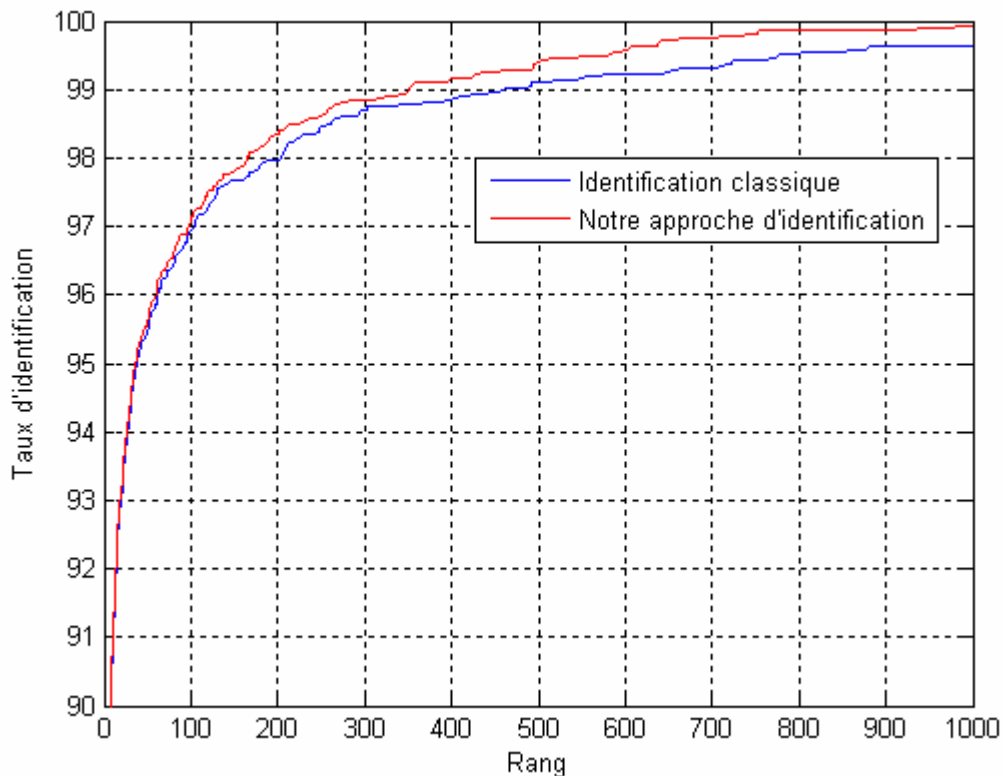


Figure 6.14 : Comparaison des courbes de scores cumulés de notre approche et celle de l'identification classique. La fusion des partitions générées par BIRCH et les K-moyennes implique le sous ensemble d'identités retenu pour l'identification finale. Celle-ci est réalisée avec les vecteurs attributs Eigenfaces ($g = 131$).

6.8.2 Temps de réponse

Afin de mesurer objectivement la complexité de notre algorithme d'identification, nous avons comparé sa vitesse d'exécution avec celle de l'identification classique. La même méthode de description des images faciales et la même mesure de similarité sont utilisées pour la phase finale d'identification des individus de notre approche et pour l'identification classique. Ceci nous procure un moyen objectif pour la comparaison de la complexité et la vitesse d'exécution de notre méthode d'identification.

L'extraction des attributs à partir des images faciales est plus ou moins coûteuse pour chaque méthode de description utilisée. Les *Eigenfaces* et les *Fisherfaces* procurent une extraction assez aisée et rapide des attributs en multipliant simplement l'image faciale par le sous-espace de visage construit. Ainsi, sur les 3111 images de l'ensemble de test dont nous disposons, cette opération nécessite en moyenne 8 secondes pour extraire un vecteur d'attributs de 131 composantes. Ceci est équivalent à une moyenne de 2,5 ms par une image de taille 64 x 64 pixels. Cependant pour extraire les moments de Zernike d'ordre 14 de tout l'ensemble de test, le temps requis est de 1280 secondes ce qui correspond à 0,43 seconde par image. Ceci compromet sensiblement la rapidité de notre démarche en employant un tel descripteur que ce soit au cours de l'étape de classification ou au cours de celle d'identification finale. Il est nécessaire donc de privilégier les méthodes d'extraction rapide des vecteurs attributs décrivant les images faciales. En effet, notre approche d'identification nécessite au moins deux vecteurs attributs par image requête : un vecteur pour la phase de classification et l'autre

pour l'identification finale. Ceci requiert deux étapes d'extraction des attributs qui doivent être performantes et aussi rapide l'une que l'autre. Par exemple, l'utilisation des *Eigenfaces* en combinaison avec les *Fisherfaces* pour les deux étapes de classification et d'identification peut améliorer la vitesse d'exécution globale de notre approche. Cependant l'utilisation des moments de Zernike pour extraire les vecteurs attributs de classification combinés à l'une de ces dernières techniques alourdit considérablement le temps d'exécution de notre approche par rapport à l'approche classique d'identification.

Afin de déterminer précisément le temps d'identification requis de notre approche d'identification, nous avons mesuré le temps de la recherche indépendamment du temps requis pour l'extraction des attributs. Le tableau 6.15 illustre les temps nécessaires pour établir une décision d'identification pour chaque type de descripteur utilisé pour les deux phases de la recherche : la C-Identification et l'identification finale. Le temps de l'approche classique d'identification nécessaire pour effectuer la recherche de l'identité requête par rapport à toutes les images de la galerie est illustré dans la dernière colonne du tableau. Il est à rappeler que notre PC de test pour ces évaluations est doté d'un processeur Intel Pentium cadencé à 1,4 GHz et d'une mémoire vive de 1Go.

Stratégies	Temps de réponse (s)		Temps de recherche total de notre approche (s)	Temps d'exécution de l'approche classique (s)
	C-Identification	Identification		
C-Ident + Ident				
LDA + PCA (L1)	1,37	70,13	71,5	300
Zernike + PCA (L1)	7,64	61,96	69,6	
PCA + LDA (L2)	1,92	9,98	11,9	190
Zernike + LDA (L2)	7,64	8,54	16,18	
PCA + Zernike (L2)	1,92	10,88	12,8	176
LDA + Zernike (L2)	1,37	10,33	11,7	

Tableau 6.15 : Comparaison des temps de la recherche en secondes de l'approche classique et de notre approche d'identification indépendamment du temps requis pour l'extraction des attributs.

Le processus d'identification des partitions (C-Identification) est relativement de faible coût par rapport à l'identification classique. En effet, la C-Identification consiste à calculer seulement la distance des individus aux centres des classes puis à ordonner ces distances en mode croissant. D'autant plus que les vecteurs attributs utilisés pour cette phase sont d'une dimensionnalité réduite par rapport à ceux utilisés pour l'identification (cf. Tableau 5.11). Ainsi nous constatons une baisse notable du temps de calcul pour toutes les stratégies mises en œuvre.

Le temps d'exécution varie suivant plusieurs paramètres. Tout d'abord, un nombre de classes important augmente la complexité de la C-Identification. Ensuite, la taille des vecteurs attributs utilisés pour la C-Identification et l'identification finale influe la vitesse d'exécution de la recherche. Des vecteurs attributs plus volumineux nécessitent plus d'espace mémoire et plus de temps de calcul. Etant donné que la signature établie par les moments de Zernike pour la classification a une taille (200 attributs) supérieure à celles des *Eigenfaces* (16 attributs) et des *Fisherfaces* (4 attributs), le temps de C-Identification des moments de Zernike (7,64s) est bien supérieur à celui de ces deux dernières méthodes. D'autre part, les distances choisies

pour l'identification finale nécessitent plus ou moins de temps de calcul. En effet, la distance de Manhattan (L_1) utilisée avec les attributs *Eigenfaces* consomme beaucoup plus de temps de calcul que la distance euclidienne (L_2). La différence du temps d'identification des deux premières stratégies du tableau 6.15 est nettement remarquée par rapport aux suivantes. Finalement, le taux de pénétration obtenu à la suite du partitionnement et de la sélection des identités retenues pour l'identification finale implique une sensible variation du temps d'identification total. Une baisse du taux de pénétration (pourcentage des identités retenues pour l'identification finale par rapport à la taille de la galerie) implique une nette amélioration du temps d'exécution final.

Nous rappelons que le temps nécessaire pour l'extraction des attributs de toutes les images de test est approximativement de 8 secondes pour les *Eigenfaces* et les *Fisherfaces* et de 1280 secondes pour les moments de Zernike. En considérant ce temps de calcul, on remarque qu'une combinaison des *Eigenfaces* et des *Fisherfaces* respectivement pour la C-Identification et l'identification finale conserve une nette amélioration du temps d'identification de notre approche. Ainsi, on obtient 25 secondes pour la stratégie « PCA + LDA (L_2) » contre 190 secondes pour l'approche classique d'identification.

Finalement, nous avons mesuré le temps d'identification en effectuant la combinaison entre le classifieur basé sur les attributs *Fisherfaces* avec la méthode de classification BIRCH et le classifieur basé sur les moments de Zernike avec une classification par les K-moyennes. Cette stratégie de combinaison de deux classifieurs pour l'étape de partitionnement de notre base de données est suivie par une identification avec les attributs *Eigenfaces* (voir tableau 6.14). La combinaison des deux classifieur, étant consommatrice de temps de calcul, a augmenté le temps d'identification total de notre approche à 170 secondes. Certes la complexité est plus importante qu'avec une classification simple, mais nous gardons toujours un temps de calcul inférieur au 300 secondes nécessaire pour l'identification classique avec les vecteurs *Eigenfaces*.

6.9 Conclusion

Nous avons proposé dans ce chapitre, un cadre plus générique en termes de données et en termes de méthodes de classification pour notre approche d'identification. Outre les K-moyennes, nous avons exploré dans ce chapitre la méthode des centres mobiles de Forgy et celle de BIRCH. Ces différentes méthodes de clustering sont appliquées dans le but de partitionner l'espace de visage en des classes homogènes et bien séparées. Les trois schémas de description du contenu des images faciales par les attributs *Eigenfaces*, *Fisherfaces* et moments de Zernike sont développés pour aboutir à des signatures variées des identités de la galerie. Après l'analyse de la réduction obtenue par ces méthodes de représentation et de clustering, nous avons proposé une stratégie de combinaison de classifieurs permettant d'obtenir un espace plus réduit pour l'identification finale. Ces classifieurs diffèrent aussi bien par leurs données d'entrées à travers les différentes représentations que par les méthodes classificatoires mises en place. Ainsi, nous avons exploré un large spectre de possibilités pour optimiser chacune des étapes de notre approche d'identification et découvrir ses limites de fonctionnement.

En validant notre approche sur un corpus de données de grande taille et assez varié le rendant plus réaliste, nous avons dégagé une amélioration considérable du temps de calcul tout en assurant des résultats similaires voire incluant une légère amélioration de l'identification finale. Au pire des cas, nous avons un temps réduit de presque la moitié. Plusieurs

enseignements peuvent être tirés pour minimiser ce temps de calcul comme le choix d'une méthode d'extraction rapide des attributs et la préférence de la distance Euclidienne L2 qui est beaucoup plus rapides que la distance de Manhattan L1.

Conclusion Générale

Cette thèse s'inscrit dans le contexte de l'identification automatique de visages dans des images numériques. Nous avons introduit une approche d'identification mettant en œuvre une classification non supervisée pour réduire la taille de la base des identités avec lesquelles sont appariés les visages requêtes. L'utilisation conjointe de différents classifieurs permet d'améliorer les performances, tant au niveau de la précision qu'au niveau de la vitesse d'exécution. Dans cette section, nous rappelons dans un premier temps les apports de cette thèse, avant de discuter certaines perspectives de ce travail.

Localisation du visage.

Nous avons mené au cours de cette thèse une étude sur la localisation du visage dans une image faciale. Cette étape qui s'inscrit dans la chaîne de reconnaissance automatique faciale est très importante pour normaliser le visage et initier la reconnaissance proprement dite. Nous avons pu voir, à travers la présentation des deux méthodes développées au laboratoire (moments de Zernike par M. Saaidia et *Eigenfaces* par nous-même) que la détection parfaite du visage était un objectif difficile à atteindre. La localisation de 30 points d'un contour de visage par un réseau de neurones entraîné par l'une ou l'autre des méthodes est à l'avantage des moments de Zernike pour élaborer le vecteur d'entrée du réseau. Le visage une fois détecté, une procédure classique de normalisation est appliquée. Meilleure est la précision de détection du visage, plus grande sont les performances des systèmes de reconnaissance qui s'ensuivent.

Le projet IV² nous a permis de travailler sur plusieurs aspects techniques pour maintenir et paramétrer les différents capteurs ainsi que pour contrôler l'éclairage d'une manière automatique suivant un protocole bien défini. Au terme des campagnes d'acquisition, un travail énorme a été effectué par moi-même et d'autres partenaires du projet pour vérifier, corriger et organiser les données acquises afin de préparer une version exploitable de la base. D'autre part, ma participation à la première campagne d'évaluation sur les données IV² m'a permis de comprendre plusieurs problématiques et défis liés non seulement à la reconnaissance des visages 2D mais aussi à la reconnaissance des visages 3D, de l'iris et de la voix. Bien que l'évaluation des systèmes d'identification n'ait pas été retenue par l'ensemble des partenaires, ce projet nous a permis de valider le bon fonctionnement des algorithmes *Eigenfaces* (PCA) et *Fisherfaces* (LDA) qui ont été utilisés plus tard dans la mise en œuvre de notre approche d'identification. Cette décision des partenaires du projet témoigne encore une fois de la complexité de la tâche de l'identification qui consiste à comparer l'ensemble des images de test avec l'ensemble des images de la galerie.

Corpus de données.

Un travail non négligeable pour l'acquisition d'autres bases Benchmark (XM2VTS, FERET, AR, ORL et CVL) et leur mise en exploitation a été effectuée durant le projet IV². Les 300 identités de la base IV², ajoutées aux données provenant de ces bases de données nous ont permis de constituer un large corpus de travail de 1880 individus. Ce corpus nous a permis d'évaluer rigoureusement les performances d'un système d'identification dans les bases de données faciales de grandes tailles.

Notre approche d'identification.

Nous avons exposé dans les deux derniers chapitres de cette thèse une approche originale de réduction de l'espace de recherche par classification non supervisée. A partir des trois méthodes d'extraction d'attributs retenues, nous avons mis en œuvre une approche de partitionnement des données. Le principe de base de cette stratégie originale consiste à conduire l'identification en deux temps. Dans un premier temps on recherche, à partir de l'identité requête, un sous-ensemble de la base de données présentant une forte probabilité de ressemblance avec l'identité requête. Dans un deuxième temps, l'identification proprement dite ne sera conduite que dans cet ensemble réduit de visages. Une analyse fine des classes d'images de visages obtenues a permis d'illustrer l'intérêt de cette approche de classification non supervisée.

Notre approche de partitionnement a été mise en œuvre non seulement avec les trois classes de description des données faciales mais aussi à travers différentes méthodes de classification non supervisée. En effet, outre les K-moyennes, nous avons exploré la méthode des centres mobiles de Forgy et celle de BIRCH afin de choisir la meilleure base d'attributs et la meilleure méthode de clustering pour partitionner les identités de la galerie. En disposant des trois partitions de la galerie, les meilleures au vu des trois classes de descripteurs en termes des taux de réduction engendrés, nous avons établi l'identification finale des identités sur les clusters retenus par chacun des différents classifieurs. Ainsi, nous avons exploré un large spectre de combinaison et de possibilités pour optimiser chacune des étapes de notre approche d'identification et pour découvrir les limites de son fonctionnement.

Après avoir analysé la réduction obtenue par chaque méthode, nous avons proposé une stratégie de combinaison des classes optimales retenues par chacun des classifieurs. Cette étape a permis d'améliorer le taux de pénétration (de réduction) de l'espace de recherche. Nous avons réalisé un gain sur notre grand corpus de travail en réduisant la taille de la galerie à 50% et en maintenant un taux d'erreur de la recherche quasi nul.

Nous avons comparé finalement l'approche d'identification directe et notre approche d'identification incluant l'étape de classification non-supervisée. Bien que les taux d'identification soient restés quasi identiques, une amélioration des scores cumulés a été mise en évidence suite à l'identification finale effectuée sur les classes et les clusters fusionnés. Nous avons montré aussi les excellentes performances en temps de calcul de notre approche avec des méthodes d'extraction rapide des attributs tels que les *Eigenfaces* et les *Fisherfaces*.

Perspectives.

Les perspectives de ce travail sont nombreuses. Dans un premier temps, nous souhaitons développer d'autres méthodes de description des images faciales plus rapides. Dans ce travail, nous avons mis en œuvre trois méthodes à savoir les *Eigenfaces*, les *Fisherfaces* et les moments de Zernike. Nous pensons choisir des méthodes alternatives qui exploitent des informations différentes des images de visages pour augmenter l'indépendance des classifieurs avec l'identification finale.

L'indépendance totale des descripteurs et des signatures utilisés pour chaque individu peut améliorer considérablement le taux d'identification final. En effet, la réduction de la moitié de l'espace de recherche peut diminuer le taux d'erreur de l'identification à sa moitié dans ces conditions. C'est dans cette perspective que d'autres modalités biométriques que le visage 2D peuvent être utilisées et amélioreront très certainement la précision de l'identification. Pour

cela nous souhaitons inclure la base de données FRGC dans notre corpus de travail. Cette base de données avec la base IV² et XM2VTS disposent, outre des images faciales 2D, de modèles 3D de visages. Une évaluation de l'identification sur la base de différents descripteurs à partir de ces deux modalités peut être une intéressante alternative. D'autre part, les étapes de la classification non supervisée et de l'identification finale peuvent s'appuyer sur une combinaison d'autres caractéristiques extraites à partir des empreintes digitales, de l'iris, de la forme de la main, ... Notre approche d'identification proposée permet d'inclure de telles caractéristiques et d'élaborer une identification plus efficace pour les applications manipulant de grandes bases de données biométriques.

Enfin, la mise en œuvre de notre approche de partitionnement pour les systèmes d'identification en monde ouvert est nécessaire pour la plupart des applications de sécurité. Dans ce cadre, l'étape de la classification non supervisée et du partitionnement de la base de données préenregistrée peut améliorer aussi les performances de détection en élagant les identités qui engendrent une fausse alarme. En effet, on ne risque pas un faux rejet lorsque l'utilisateur n'est pas enregistré dans la base de données du système. Dans ce cas, plus on réduit la galerie par le module de classification, plus le système sera capable de juger que l'utilisateur est inconnu par rapport à une base d'individus réduite.

Références

- [AB05] S. Avidan, M. Butman. "*The Power of Feature Clustering: An Application to Object Detection*", Advances in Neural Information Processing Systems, vol. 17, pp. 57-64, 2005.
- [Abb04] Bouchra Nakad Abboud. "Analyse d'expressions faciales par modèles d'apparence", Thèse de doctorat, Université de Technologie de Compiègne, France, 2004.
- [ABR64] M. Aizerman, E. Braverman, L. Rozonoer. "*Theoretical foundations of the potential function method in pattern recognition learning*", Automation and Remote Control, vol. 25, pp. 821-837, 1964.
- [AC06] O. Arandjelovic, R. Cipolla. "*Face recognition from video using the generic shape-illumination manifold*", the 9th European Conference on Computer Vision, pp.7-13, Autriche, 2006.
- [AEBN05] G. Amayeh, A. Erol, G. Bebis, M. Nicolescu. "*Accurate and efficient computation of high order zernike moments*", First International Symposium, ISVC, pp. 462-469, 2005.
- [AG05] C. B. Abdelkader, P. Griffin. "*A Local Region-Based Approach to Gender Classification from Face Images*", Proc. 2005 IEEE CS Conf. Computer Vision and Pattern Recognition, p. 52, 2005.
- [Ali05] S. Alirezaee, M. Ahmadi, H. Aghaeinia, K. Faez. "*A weighted pseudo-zernike feature extractor for face recognition*", proceedings of 11th IEEE International Conference on Systems, Man, and Cybernetics, vol. 3, pp. 2128-2132, USA, 2005.
- [AMM08] A. Albiol, D. Monzo, A. Martin, J. Sastre, A. Albiol, "*Face recognition using HOG-EBGM*", Pattern Recognition Letters, vol. 29, no. 10, pp. 1537-1543, July 2008.
- [Amo06] B. B. Amor. "Contributions à la modélisation et à la reconnaissance faciales 3D", Thèse de Doctorat, Ecole Centrale de Lyon, France, 2006.
- [BAS96] S. O. Belkasim, M. Ahmadi, and M. Shridhar. "*Efficient algorithm for fast computation of Zernike moments*", IEEE 39th Midwest Symp. on Circuit and Systems, vol. 3, pp. 1401-1404, 1996.
- [BBD93] A. M. Burton, V. Bruce, N. Dench. "*What's the difference between men and women? Evidence from facial measurement*", Perception 22, pp.153-176, 1993.
- [BBP01] D.M. Blackburn, J.M. Bone, and P.J. Phillips. "*Facial Recognition Vendor Test 2000: Evaluation Report*", Technical Report A269514, National Institute of Standards and Technology, 70 pages, 2001.
- [BC07] M. Ben Ahmed, M. Charrad. "*Tutorial des Méthodes Classificatoires*", Ecole Nationale des Sciences de l'Informatique, Université de la Manouba, Tunisie, 84 pages, 2007.
- [BDDBS02] K. Back, B. A. Draper, J. R. Beveridge, K. She. "*PCA vs ICA: A comparison on the FERET data set*", proceedings of the IEEE international Conference on Computer Vision and Pattern Recognition, pp. 824-827, 2002.
- [BFR98] P. S. Bradley, U. Fayyad, C. Reina. "*Scaling Clustering Algorithms to Large Databases*", proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining. pp. 9-15, New York, August, 1998.

- [BGV92] B. E. Boser, I. M. Guyon, V. N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers", In Fifth Annual Workshop on Computational Learning Theory, pp.144-152, ACM, 1992.
- [BHB98] V. Bruce, P.J.B. Hancock, A.M. Burton, "Human Face Perception and Identification", Face recognition : from theory to applications, vol. 163, pp. 51-72, 1998.
- [BHK97] P. Belhumeur, J. Hespanha, D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, pp. 711-720, July 1997.
- [Bio.eu] Association BioSecure, <http://biosecure.it-sudparis.eu/AB/>.
- [BJ02] F. Bach, M. I. Jordan, "Kernel independent component analysis", Journal of Machine Learning Research, vol. 3, pp.1-48, 2002.
- [BLFM03] M.S. Bartlett, G. Littlewort, I. Fasel, J.R. Movellan, "Real Time Recognition of Facial Expressions: Development and Applications to Human Computer Interaction", International Conference on Computer Vision and Pattern Recognition (CVPR), vol.5, pp. 53-58, 2003.
- [BLS98] M.S. Bartlett, H.M. Lades, T. SEJNOWSKI, "Independent component representation for face recognition", Proceedings SPIE Symposium on Electronic Imaging: Science and Technology, vol. 3299, pp. 528-539, 1998.
- [BP93] R. Brunelli, T. Poggio, "Face Recognition : Features versus Templates", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.15, n°.10, pp.1042-1052, 1993.
- [BR07] S. Baluja, H.A. Rowley, "Boosting Sex Identification Performance", Int'l J. Computer Vision, vol. 71, no. 1, pp. 111-119, 2007.
- [BRP04] R. M. Bolle, N. K. Ratha, S. Pankanti, "Error analysis of pattern recognition systems - the subsets bootstrap", Computer Vision and Image Understanding, vol. 93, n°1, pp. 1-33, January 2004.
- [BW54] Bhatia, A. and Wolf, E. "On the circle polynomials of Zernike and related orthogonal sets". Proc. Cambridge Philosophical Society, vol. 50, pp. 40-48, 1954.
- [CBA04] N.P. Costen, M. Brown, S. Akamatsu, "Sparse models for gender classification", IEEE Proceeding in Automatic Face and Gesture Recognition, pp. 201- 206, May 2004.
- [CBL08] A. Chaari, M. Ben Ahmed, S. Lelandais, "Appearance based face identification and classification. A Combining approach", International Conference on Information and Communication Technologies: From Theory to Applications, pp. 1-6, Tunisia, 2008.
- [CC04] D. Cristinacce, T. Cootes, "A comparison of shape constrained facial feature detectors", Proceedings of the 6th International Conference on Automatic Face and Gesture Recognition, pp. 375-380, 2004.
- [CCZ05] R. Chellappa, A. Roy-Chowdhury, S. Zhou, "**Recognition of Humans and Their Activities Using Video**", Morgan & Claypool Publishers, 174 pages, 2005.
- [CET01] T.F. Cootes, G.J. Edwards, C.J. Taylor, "Active Appearance Models", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.23, n°.6, pp. 681-685, 2001.
- [CF90] G. Cottrel, M. Flemming, "Face Recognition using unsupervised feature extraction", Proceedings of the International Neural Network Conference, vol. 1, pp.322-325, 1990.

- [CH97] A. J. Colmenarez, T. S. Huang, "*Face detection with information based maximum discrimination*", IEEE International Conference on Computer Vision and Pattern Recognition, pp.782-787, 1997.
- [CH07] J. Colineau, J. d'Hose. "*Techno Vision IV² - Evaluation d'algorithmes biométriques - Iris, Visage 2D, Visage 3D, Multimodalité*", Rapport sur les évaluations du projet IV², 2007.
- [Che05] D. Cherif. "Techniques d'extraction des connaissances appliquées aux données du Web", Mémoire de Mastère en Informatique, Ecole Nationale des Sciences de l'Informatique de Tunis, Laboratoire RIADI, 2005.
- [Chr04] C. Christophe. "SearchXQ : une méthode d'aide à la navigation fondée sur Ω -means, algorithme de classification non-supervisée. Application sur un corpus juridique français". Thèse de doctorat. Ecole des Mines Paris, Décembre 2004.
- [CL08] J.T. Chien; C.P. Liao, "*Maximum Confidence Hidden Markov Modeling for Face Recognition*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.30, n° 4, pp.606-616, 2008.
- [CLB07] A. Chaari, S. Lelandais, M. Ben Ahmed, "*Face classification scheme simplifying identification in large scale biometric databases*", Fourth IEEE International Conference on Systems, Signals & Devices, March 2007.
- [CLB08] A. Chaari, S. Lelandais, M. Ben Ahmed, "*A New Clustering Approach for Face Identification*", First Workshops on Image Processing Theory, Tools and Applications, pp. 1-6, 2008.
- [CLS⁺08] A. Chaari, S. Lelandais, M. Saaidia, V. Vigneron, M. Ben Ahmed, "*Comparaison d'approches pour la détection et l'identification de visages 2D*", 14ème Colloque National de la Recherche en IUT (CNRIUT), France, 2008.
- [Clu04] Technique de Contrôle d'Accès par Biométrie, <http://www.clusif.asso.fr>, dossier technique, CLUSIF, 2004.
- [CM98] V. Cherkasski, F. Mulier, "*Learning from Data: Concepts, Theory, and Methods*", Nueva York, EUA: Wiley, 464 pages, 1998.
- [CMM00] R. Cappelli, D. Maio, D. Maltoni, "*Indexing Fingerprint. Databases for Efficient 1:N Matching*", in proc. International Conference on Control, Automation, Robotics and Vision, Singapore (ICARCV), 2000.
- [CMMN03] R. Cappelli, D. Maio, D. Maltoni, L. Nanni, "*A two-stage fingerprint classification system*", in proceedings ACM SIGMM Multimedia Biometrics Methods and Applications Workshop (WBMA03), Berkley, pp.95-99, November 2003.
- [Con04] D. Conte, P. Foggia, C. Sansone, M. Vento, "*Thirty years of graph matching in pattern recognition*", Int. Journal of Pattern Recognition and Artificial Intelligence, vol. 18, n°3, pp.265-298, 2004.
- [CVL.si] CVL Face Database, <http://www.lrv.fri.uni-lj.si/facedb.html>.
- [Dan07] Thanh-Ha Dang, "Mesures de discrimination et leurs applications en apprentissage inductif", Thèse de doctorat, Université Paris 6, soutenue le 10/07/2007.
- [Dau07] J. Daugman, "*New Methods in Iris Recognition*", IEEE Transactions on Systems, Man, and Cybernetics, Vol. 37, No. 5, 2007.
- [DBBB03] B. Draper, K. Baek, M.S. Bartlett, and R. Beveridge, "*Recognizing Faces with PCA and ICA*", Computer Vision and Image Understanding, vol. 91, pp. 115-137, 2003.

- [DGB08] K. Delac, M. Grgic, M.S. Bartlett, "**Recent Advances in Face Recognition**", Publisher: IN-TECH, Autriche, 236 pages, December 2008.
- [Did82] E. Diday, J. Lemaire, J. Pouget et F. Testu, "*Eléments d'analyse des données*", Dunod Informatique, 1982.
- [EC97] K. Etemad, R. Chellappa, "*Discriminant Analysis for Recognition of Human Face Images*", Journal of the Optical Society of America, vol.14, pp.1724–1733, 1997.
- [ESSEX] VISION GROUP OF ESSEX UNIVERSITY – Face database <http://cswww.essex.ac.uk/allfaces/index.html>.
- [Fau05] M. Faundez-Zanuy, "*Signature recognition state-of-the-art*", IEEE Aerospace and Electronic Systems Magazine, Volume 20, Issue 7, pages: 28 – 32, July 2005.
- [FBVC01] R. Feraud, O. J. Bernier, J-E. Viallet, M. Collobert, "*A Fast and Accurate Face Detection Based on Neural Network*", IEEE Trans. Pattern Analysis and Machine Intelligence, 2001.
- [Fis07] J. Fiscus, J. Ajot, J. Garofolo, G. Doddington, "*Results of the 2006 Spoken Term Detection Evaluation*", in Proc. 2007 SIGIR Workshop on Searching Spontaneous Conversational Speech, Amsterdam, July 2007.
- [FJ08] J. Feng, A. K. Jain, "*Filtering Large Fingerprint Database for Latent Matching*", the 19th International Conference on Pattern Recognition, pp. 1-4, Florida, USA, 2008.
- [FK02] B. Fröba, C. Küblbeck, "*Robust face detection at video frame rate based on edge orientation features*", in: Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition AFGR, pp. 342–347, 2002.
- [FL03] B. Fasel, J. Luetttin, "*Automatic Facial Expression Analysis: A Survey*", Pattern Recognition, 36:259-275, 2003.
- [For65] E. W. Forgy, "*Cluster analysis of multivariate data: efficiency versus interpretability of classifications*", Biometrics, 21, pp.768-769, 1965.
- [FRH.org] The Face Recognition Homepage, www.face-rec.org/interesting-papers.
- [FS95] Y. Freund, R. E. Schapire. "*A decision-theoretic generalization of on-line learning and an application to boosting*". In Computational Learning Theory: Eurocolt '95, Springer-Verlag, pp. 23-37, 1995.
- [FT86] O.D. Faugeras, G. Toscani, "*The Calibration Problem for Stereo*", Proceedings of the CVPR, pp.15-20 1986.
- [Fuk90] K. Fukunaga, "**Introduction to statistical pattern recognition**", (2nd ed.), Academic Press Professional, pages: 592, 1990.
- [GAP+02] H. Gupta, A.K. Agrawal, T. Pruthi, C. Shekhar, R. Chellappa, "*An Experimental Evaluation of Linear and Kernel-Based Methods for Face Recognition*", In Proc. Of the IEEE Workshop on Applications of Computer Vision, pp. 13-19, 2002.
- [GBMK04] R. Gross, S. Baker, I. Matthews, T. Kanade, "*Face Recognition Across Pose and Illumination*", Handbook of Face Recognition, chapitre 9, Stan Z. Li and Anil K. Jain, ed., Springer-Verlag, 27 pages, 2004.
- [GC88] A.D. Gordon, A. De, "*Cata. Stability and influence in sum of squares clustering*", Metron. vol. 46, pp. 347-360, 1988.

- [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), pp.1408-1423, 2004.
- [GL00] I. Gauthier, N. K. Logothetis, "*Is Face Recognition Not so Unique After All ?*", J. of Cognitive Neuropsychology, Vol. 17, No 1-3, pp. 125-142, 2000.
- [GL02] Y. Gao; M.K.H. Leung, "*Face recognition using line edge map*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24, n°.6, pp.764 -779, 2002.
- [GMM97] M. Golfarelli, D. Maio, D. Maltoni, "*On the error-Reject Tradeoff in Biometric Verification Systems*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 19, n° 7, pp. 786-796, 1997.
- [GMP00] S. Gong, S.J. McKenna, A. Psarrou, "**Dynamic Vision: From Images to Face Recognition**", Imperial College Press, 364 pages, Londres, Angleterre, 2000.
- [GMP03] P.J. Grother, R.J. Micheals, P. J. Phillips, "*Face Recognition Vendor Test 2002 Performance Metrics*", Proceedings 4th International Conference on Audio Visual Based Person Authentication (AVBPA), pp. 937-945, 2003.
- [Gro04] R. Gross, "*Face Databases*", Handbook of face Recognition, chapitre 13, Springer-Verlag, pp. 301-328, 2004.
- [GS06] Y. Gizatdinova, V. Surakka, "*Feature-Based Detection of Facial Landmarks from Neutral and Expressive Facial Images*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.28, no.1, pp. 135-139, 2006.
- [GSC01] R. Gross, J. Shi, J.F. Cohen, "*Quo vadis Face Recognition?*", in: Proceedings of Third Workshop on Empirical Evaluation Methods in Computer Vision, 2001.
- [Gue08] S. Guerfi, "*Authentification d'individus par reconnaissance de caractéristiques biométriques liées au visages 2D/3D*", Thèse de doctorat, Université d'Evry, 2008.
- [Gut01] S. Gutta, J. Huang, C. J. Liu, H. Wechsler, "*Comparative Performance Evaluation of Gray Scale and Color Information for Face Recognition Tasks*", 3rd Int. Conf. On Audio and Video-Based Biometric Person Authentication (AVBPA), pp. 38-43, 2001.
- [GSTL02] J. Gu, H. Z. Shua, C. Toumoulinb, L. M. Luo, "*A novel algorithm for fast computation of Zernike moments*", Pattern Recognition, vol 35, pp. 2905-2911, 2002.
- [GZ07] J. Gu, J. Zhou, "*Analysis of Singular Points in Fingerprints based on Topological Structure and Orientation Field*", International Conference on Biometrics, pp. 261-270, 2007.
- [GZT00] C. Garcia, G. Zikos, G. Tziritas, "*Wavelet Packet Analysis for Face Recognition*", Image and Vision Computing, Vol.18(4), pp.289--297, February 2000.
- [HAF03] J. Haddadnia, M. Ahmadi, K Faez, "*An Efficient Feature Extraction Method with Pseudo Zernike Moment in RBF Neural Network-Based Human Face Recognition System*", EURASIP, JASP, vol. 9, pp. 890-891, 2003.
- [HB95] A. J. Howell, H. Buxton, "*Invariance in radial basis function neural networks in human face classification*", Neural Processing Letters, Vol. 2, n° 3, pp. 26-30, 1995.
- [HB96] H. Hill, V. Bruce, "*Effect of Lighting on Matching Facial Surfaces*", Journal of Experimental Psychology: Human Perception and Performance, 22, pp.986-1004, 1996.
- [HHB03] J. Huang, B. Heisele, and V. Blanz, "*Component-Based Face. Recognition with 3D Morphable Models*", In Proceedings, International Conference on Audio- and Video-Based Person Authentication, pp.27-34, 2003.

- [HHP01] B. Heisele, P. Ho, and T. Poggio, "*Face recognition with support vector machines: global versus component-based approach*", In Proc. 8th International Conference on Computer Vision, vol.2, pp.688-694, 2001.
- [Hje01] Erik Hjelm, "*Face Detection : A Survey*", Computation Vision and Image Understanding, Vol 83, 236-274, 2001.
- [Hua98] R.-J. Huang, "*Detection strategies for face recognition using learning and evolution*", Ph.D. thesis, George Mason University, Fairfax, Virginia, 1998.
- [IV².fr] Le site web du projet IV²: <http://iv2.ibisc.fr/>.
- [JBP99] A.K. Jain, R. Bolle, S. Pankanti, "*Biometrics: Personal Identification in Networked Society*", Springer, 411 pages, 1999.
- [JHC92] A. Johnston, H. Hill, N. Carman, "*Recognizing Faces: Effects of Lighting Direction*", Inversion and Brightness Reversal, Cognition, vol. 40, pp. 1-19, 1992.
- [JHP00] A.K. Jain, L. Hong, S. Pankanti, "*Biometrics: Promising frontiers for emerging identification market*", ACM 43, 2, pp. 91-98, 2000.
- [JKF01] O. Jesorsky, K. Kirchberg, R. Frischholz, "*Robust face detection using the hausdorff distance*", proceedings of the 3rd International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA), pp. 90–95, 2001.
- [JKL03] J. Jang, K. Kim, Y. Lee, "*Efficient Algorithm of Eye Image Check for Robust Iris Recognition System*", Computer Analysis of Images and Patterns, Springer, pp. 301-308, 2003.
- [JKLM00] K. Jonsson, J. Kittler, Y. Li, J. Matas, "*Learning Support Vectors for Face Verification and Recognition*", proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, pp. 208-213, 2000.
- [JNR05] A. Jain, K. Nandakumara, A. Ross, "*Score normalization in multimodal biometric systems*", Pattern Recognition Society, Elsevier, pp. 2270 - 2285, 2005.
- [JP00] A. K. Jain, S. Pankanti, "*Fingerprint Classification and Matching*", Handbook for Image and Video Processing, A. Bovik (ed.), Academic Press, 2000.
- [JPP⁺04] A. K. Jain, S. Pankanti, S. Prabhakar, L. Hong, A. Ross, "*Biometrics: A Grand Challenge*", Proceedings of International Conference on Pattern Recognition, Vol.2, pp. 935-942, UK, 2004.
- [JRP04] A. K. Jain, A. Ross, S. Prabhakar, "*An Introduction to Biometric Recognition*", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No. 1, 2004.
- [KAG03] N. Kumar, V. Abhishek, G. Gautam, "*A novel approach for person authentication and content-based tracking in videos using kernel methods and active appearance models*", IEEE International Conference on Systems, Man and Cybernetics, Vol.2, pp.1384-1389, 2003.
- [Kan77] T. Kanade, "*Computer recognition of human faces*", Interdisciplinary Systems Research, Vol. 47, 1977.
- [Kel70] M.D. Kelly, "*Visual identification of people by computer*", Technical report AI-130, Stanford AI Project, 1970.
- [KFG08] H. R. Kanan, K. Faez, Y. Gao, "*Face Recognition Using Adaptively Weighted Patch PZM Array from a Single Exemplar Image per Person*", Pattern Recognition, Vol. 41, No. 12, pp.3799-3812, 2008.

- [Kir00] M. Kirby, "*Dimensionality Reduction and Pattern Analysis : An Empirical Approach*", Wiley, New York, 2000.
- [Kra91] M. A. Kramer, "*Nonlinear principal component analysis using autoassociative neural networks*", *AICHE Journal*, vol. 37, no2, pp. 233-243, 1991.
- [KS90] M. Kirby, 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, Jan. 1990.
- [Lad93] M. Lades, C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. Wurtz, and W. Konen, "*Distortion invariant object recognition in the dynamic link architecture*", *IEEE Trans. Computers*, 42(3):300–311, March 1993.
- [Law97] S. Lawrence, C. L. Giles, A. Tsoi, A. Back, "*Face recognition: A convolutional neural-network approach*", *IEEE Trans. on Neural Networks*, vol. 8, n° 1, pp. 98-113, 1997.
- [LC03] X. Liu, T. Chen, "*Video-based face recognition using adaptive hidden Markov models*", *CVPR*, vol. 1, pp.340-345, 2003.
- [LCJ06] X. Lu, H. Chen, A. K. Jain, "*Multimodal Facial Gender and Ethnicity Identification*", *International Conference on Biometrics (ICB)*, pp.554-561, Hong Kong, 2006.
- [LGL01] Y. Li, S. Gong, H. Liddell, "*Modelling faces dynamically across views and over time*", *ICCV*, 1, pp.554-559, 2001.
- [Liu03] C. Liu: "*A Bayesian Discriminating Features Method for Face Detection*", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 725-740, 2003.
- [LJ04] S. Z. Li, A. K. Jain, "*Handbook of Face Recognition*", Springer, 2004.
- [LK05] K. Lee, D. Kriegman. "*Online learning of probabilistic appearance manifolds for video-based recognition and tracking*", *CVPR*, 1, pp.852-859, 2005.
- [LKP03] R. Lienhart, A. Kuranov, V. Pisarevsky, "*Empirical analysis of detection cascades of boosted classifiers for rapid object detection*", *Proceedings of the 25th DAGM-Symposium*, Magdeburg, Germany, pp. 297–304, 2003.
- [LL06] H.-C. Lian, B.-L. Lu, "*Multi-View Gender Classification Using Local Binary Patterns and Support Vector Machines*", *Proc. Third Int'l Symp. Neural Networks*, vol. 2, pp. 202-209, 2006.
- [LLT08] J. Li, X. Li, D. Tao, "*KPCA for semantic object extraction in images*", *Pattern Recognition*, vol. 41, no. 10, pp. 3244-3250, 2008.
- [LM98] J. Luettin, G. Maître, "*Evaluation protocol for the extended M2VTS database (XM2VTSDB)*", in *IDIAP Communication 98-05*, IDIAP, Switzerland, 1998.
- [LW98] C. Liu, H. Wechsler, "*Enhanced fisher linear discriminant models for face recognition*", In *Proc. the 14th International Conference on Pattern Recognition*, 1998.
- [LW00] C. Liu, H. Wechsler, "*Evolutionary pursuit and its application to face recognition*", *IEEE Trans. Patt. Anal. Mach. Intell.* 22, 570–582, 2000.
- [Mac67] J. B. MacQueen, "*Some methods for classification analysis of multivariate observations*", *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability I*, pp. 281-297, CA: University of California Press, 1967.

- [Mai02] D. Maio, D. Maltoni, R. Cappelli, JL Wayman, AK Jain, "*FVC2002: Second Fingerprint Verification Competition*", 16th International Conference on Pattern Recognition, Vol. 3, pp. 811-814, 2002.
- [MB98] A.M. Martinez, R. Benavente, "*The AR face database*", Technical Report 24, Computer Vision Center, Espagne, 1998.
- [MC96] G.W. Milligan, R. Cheng, "*Measuring the influence of individual data points in a cluster analysis*", Journal of Classification, vol. 46, pp.187-189, 1996.
- [MCG05] A. Mhatre, S. Chikkerur, V. Govindaraju, "*Indexing Biometric Databases using Pyramid Technique*", Audio and Video-based Biometric Person Authentication (AVBPA), pp. 841-849, USA, July-2005.
- [MCG⁺09] A. Mellakh, A. Chaari, S. Guerfi, J. Dhose, J. Colineau, Sylvie Lelandais, D. P.-Delacretaz, B. Dorizzi, "*2D Face Recognition in the IV2 Evaluation Campaign*", Advanced Concepts for Intelligent Vision Systems (ACIVS), Bordeaux, France, 2009.
- [MDD06] M. A. Mellakh, D. P. Delacretaz, B. Dorizzi, "*Using signal residual information of Eigenfaces for PCA face space dimensionality characteristics*", ICPR, Hong Kong, pp.574-577, 2006.
- [Mes99] K. Messer, J. Matas, J. Kittler, J. Luetttin, G. Maitre, ' "*XM2VTSDB: The Extended M2VTS Database*", Proc. AVBPA'99 Washington DC, 1999.
- [MH95] O. Monga, R. Horaud, "**Vision par Ordinateur : Outils Fundamentaux**", Editions Hermés, 425 pages, Paris, 1995.
- [MKCK01] A.J. Mansfield, G. Kelly, D. Chandler, J. Kane, "*Biometric Product Testing*", Final Report, Issue 1.0, March 2001.
- [MMJP03] D. Maltoni, D. Maio, A.K. Jain, and S. Prabhakar, "*Multimodal Biometric Systems*", Handbook of Fingerprint Recognition, Chapitre 7, Springer, 2003.
- [MMJP05] D. Maltoni, D. Maio, A.K. Jain, S. Prabhakar, "**Handbook of Fingerprint Recognition**", Springer, 340 pages, 2005.
- [Mog02] B. Moghaddam, "*Principal manifolds and probabilistic subspace for visual recognition*", IEEE Trans. Pattern Anal. Mach. Intell. 24 (6), pp. 780–788, 2002.
- [MP97] B. Moghaddam, A. Pentland, "*Probabilistic visual learning for object representation*", IEEE Trans. Pattern Analysis and Machine Intelligence, 19, 696–710, 1997.
- [MP98] H. Moon, J. Phillips, "*Analysis of PCA-Based Face Recognition Algorithms*", Empirical Evaluation Techniques in Computer Vision, pp. 835 – 855, IEEE Computer Society Press, 1998.
- [MPCG05] A. Mhatre , S. Palla , S. Chikkerur, V. Govindaraju, "*Efficient Search and Retrieval in Biometric Databases*", SPIE Defense and Security Symposium, Vol - 5779, pp. 265-273, March-2005.
- [MR08] E. Mäkinen, R. Raisamo, "*Evaluation of Gender Classification Methods with Automatically Detected and Aligned Faces*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, n° 3, pp. 541-547, 2008.
- [MR95] R. Mukundan, K.R. Ramakrishnan, "Fast computation of Legendre and Zernike moments", Pattern Recognition, vol. 28, No. 9, pp. 1433-1442, 1995.
- [MRW⁺99] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, K.R. Müller, "*Fisher discriminant analysis with kernels*", Neural Networks for Signal Processing, pp.41-48, 1999.

- [MY00] B. Moghaddam, M.-H. Yang, "*Gender Classification with Support Vector Machines*", Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 306-311, Grenoble, France, 2000.
- [NB97] H.V. Neto, D.L. Borges, "*Fingerprint classification with neural networks*", 4th Brazilian Symposium on Neural Networks (SBRN), pp. 66-72, Brésil, 1997.
- [Nist.gov] Le site web de la base de données Color FERET, <http://www.nist.gov/humanid/colorferet>
- [O'DV94] A.J. O'Toole, K.A. Deffenbacher, D. Valentine, "*Structural aspects of face recognition and the other-race*", *Memory & Cognition*, 22 : 208-224, 1994.
- [OFG97] E. Osuna, R. Freund, F. Girosi, "*Training support vector machines: an application to face detection*", *Computer Vision and Pattern Recognition*, pp.130, 1997.
- [O'T04] A.J. O'Toole, "*Psychological and Neural Perspectives on Human Face Recognition*", *Handbook of Face Recognition*, Chapitre 15, Springer, 2004.
- [PG05] J.R. Price, T.F. Gee, "*Face recognition using direct, weighted linear discriminant analysis and modular subspace*", *Pattern Recognition*, vol. 38, Issue 2, pp. 209–219, 2005.
- [Phi07] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, M. Sharpe, "*FRVT 2006 and ICE 2006 large-scale results*", Technical Report NISTIR 7408, National Institute of Standards and Technology, 2007.
- [RBK98] H. Rowley, S. Baluja, T. Kanade, "*Neural Network-Based Face Detection*", *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 20, n°1, pp. 23-38, January 1998.
- [RCBM06] Y. Rodriguez, F. Cardinaux, S. Bengio, J. Mariéthoz, "*Measuring the Performance of Face Localization Systems*", *Image and Vision Computing*, vol. 24, no. 8, pp. 882-893, Aug. 2006.
- [RJ03] A. Ross, A.K. Jain, "*Information Fusion in Biometrics*", *Pattern Recognition Letter*, vol. 24, pp.2115-2125, 2003.
- [RNJ06] A. Ross, K. Nandakumar, A. K. Jain, "**Handbook of Multibiometrics**", Springer, 202 pages, 2006.
- [RRE04] C. Rosenberger, A. Rakotomamonjy, B. Emile "*Generic target recognition*", *EUSIPCO*, pp. 1613-1616, 2004.
- [Rud94] D. L. Ruderman, "*The statistics of natural images*", *Netw.: Comput. Neural Syst.* 5, 598–605, 1994.
- [RY98] D. Reisfeld, Y. Yeshurun, "*Preprocessing of face images: Detection of features and pose normalization*", *Comput. Vision Image Understanding* 71, 1998.
- [OECD05] Rapport OECD, Direction de la science, de la technologie et de l'industrie, "*Technologies fondées sur la biométrie*", www.oecd.org, Juin 2005.
- [Oka98] K. Okada, J. Steffans, T. Maurer, H. Hong, E. Elagin, H. NEVEN, C.V.D. Malsburg, "*The Bochum/USC Face Recognition System and how it fared in the FERET Phase III Test*". *Face Recognition: From Theory to Applications*, Eds. Springer-Verlag, pp. 186-205, 1998.
- [Ono03] A. Ono, "*Face recognition with Zernike moments*", *Systems and Computers*, vol. 34, No 10, 2003.

- [PA96] P.S. Penev, J.J. Atick, "*Local feature analysis: A general statistical theory for object representation*", Network: Computation in Neural Systems, vol. 7, n° 3, pp. 477-500, 1996.
- [PD02] F. Perronnin, J.-L. Dugelay, "*Introduction à la Biométrie - Authentification des Individus par Traitement Audio-Vidéo*", Revue Traitement du Signal, volume 19, n° 4, 2002.
- [PGM04] P. J. Phillips, P. Grother, R. Micheals, "*Evaluation Methods in Face Recognition*", Handbook of Face Recognition, chapitre 14, Springer, 2004.
- [PJ02] S. Prabhakar, A.K. Jain, "*Decision-level fusion in fingerprint verification*", Pattern Recognition, Vol. 35, No. 4, pp. 861-874, 2002.
- [Phi98] P. J. Phillips, "*Support vector machines applied to face recognition*", Adv. Neural Inform. Process. Syst. Vol. 11, pp. 803–809, 1998.
- [PMRR00] P. J. Philips, H. Moon, S. A. Rizvi, P. J. Rauss, "*The FERET evaluation methodology for face-recognition algorithms*", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.22, no.10, pp.1090-1100, 2000.
- [PGM⁺03] P.J. Phillips, P. Grother, R.J Micheals, D.M. Blackburn, E Tabassi, J.M. Bone, "*Face Recognition Vendor Test 2002: Evaluation Report*", Technical Report 6965, National Institute of Standards and Technology, 56 pages, 2003.
- [PM00] D. Pelleg and A. Moore. "*X-Means: Extending k-means with efficient estimation of the number of clusters*". Proceedings of ICML-2000, pp. 727-734, 2000.
- [PMS94] A Pentland, B Moghaddam, T Starner, "*View-based and modular eigenspaces for face recognition*". Proceeding, IEEE Conference on Computer Vision and Pattern Recognition, pp. 84-91, 1994.
- [PMWP00] P. J. Phillips, A. Martin, C. L. Wilson, M. Przybocki, "*An introduction to evaluating biometric systems*", IEEE Computer, Vol 33, No. 2, pp. 56-63, 2000.
- [PS00] P.S. Penev, L. Sirovich, "*The Global Dimensionality of Face Space*", Fourth IEEE International Conference Automatic Face and Gesture Recognition, pp.26-30, 2000.
- [PTN06] Y.H. Pang, A. B.J. Teoh, D. C.L. Ngo, "*A Discriminant Pseudo Zernike Moments in Face Recognition*", Journal of Research and Practice in Information Technology, vol. 38, n° 2, pp. 197-210, May 2006.
- [PVY03] S. Palanivel, B. S. Venkatesh, B Yegnanarayana, "*Real time face recognition system using autoassociative neural network models*", ASSP, pp.833-836, 2003.
- [PW05] G. Pan, Z. Wu, "*3D Face Recognition from Range Data*", International Journal of Image and Graphics, vol. 5, n°3, 2005.
- [PWHR98] P.J. Phillips, H. Wechsler, J. Huang, P.J. Rauss, "*The FERET Database and Evaluation Procedure for Face Recognition Algorithms*", Image and Vision Computing J., vol. 16, no. 5, pp. 295-306, 1998.
- [Sch97] H. Schütze and C. Silverstein. "*Projections for efficient document clustering*". Proceedings of the 20th International ACM SIGIR Conference, 1997.
- [SCL⁺07] M. Saaidia, A. Chaari, S. Lelandais, V. Vigneron, M. Bedda, "*Face localization by neural networks trained with Zernike moments and Eigenfaces feature vectors. A comparison*", IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS), London, pp. 377-382, 2007.
- [Sim82] H. A. Simon, "*The Science of the Artificial*", MIT Press, 1982.

- [SK00] H. Schneiderman, T. Kanade, "A statistical method for 3D object detection applied to faces and cars", Conference on Computer Vision and Pattern Recognition, 2000.
- [SK87] L. Sirovich, M. Kirby, "Low-dimensional procedure for the characterization of the human face", Journal of the Optical Society of America, vol. 4, n° 3, pp.519-524, 1987.
- [SL05] P. Shih, C. Liu, "Comparative assessment of Content-Based face image retrieval in Different Color Space", Automatic Video-Based Person Authentication AVBPA, 2005.
- [SLVB07] M. Saaidia, S. Lelandais, V. Vigneron, M. Bedda, "Face detection by neural network trained with Zernike moments", 6th WSEAS International Conference on Signal Processing, Robotics and Automation, pp. 36-41, 2007.
- [SM04] G. Shakhnarovich, B. Moghaddam, "Face Recognition in Subspaces", Handbook of Face Recognition, chapitre 7, pp. 283-297, Springer-Verlag, 2004.
- [SP98] Kah-kay Sung, Tomaso Poggio, "Example-based learning for view-based human face detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998.
- [SSM98] B. Schölkopf, A. Smola, K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem", Neural Computation, vol.10, pp.1299-1319, 1998.
- [SSNT06] M. C. Santana, O. D. Suárez, J. L. Navarro, M. H. Tejera, "Gender and Identity Classification for a Naive and Evolving System", Proc. Second Workshop Multimodal User Authentication (MMUA), 2006.
- [SW96] D. L. Swets, J. Weng, "Using discriminant eigenfeatures for image retrieval", IEEE Trans. Pattern Analysis and Machine Intelligence, 18, 831– 836, 1996.
- [SY94] F. Samaria, S. Young, "HMM based architecture for face identification", Image and Vision Computing, vol. 12 No 8, pp. 537–583, 1994.
- [Tak98] B. Takacs, "Comparing Face Images Using the Modified Hausdorff Distance", Pattern Recognition, vol. 31, no. 12, pp.1873-1881, 1998.
- [Tan05] X. Tan, S.C. Chen, Z.-H. Zhou, F. Zhang, "Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft kNN ensemble", IEEE Trans. Neural Networks 16 (4), pp. 875–886, 2005.
- [TB98] N.F. Troje, H.H. Bühlhoff, "How is Bilateral Symmetry of Human Faces Used for Recognition of Novel Views", Vision Research, 38:79-89, 1998.
- [TBL03] X. Tan, B. Bhanu, Y. Lin, "Fingerprint Identification: Classification vs. Indexing", IEEE Conference on Advanced Video and Signal Based Surveillance, pp.151-156, 2003.
- [Tib92] R. Tibshirani, "Principal curves revisited", Statistics and Computation, vol. 2, pp.183-190, 1992.
- [Tol01] A. S. Tolba, "Invariant gender identification", Digital Signal Processing, 11(3), pp. 222-240, Portugal, 2001.
- [TP91] M. Turk, A. Pentland, "Eigenfaces for recognition", Journal of cognitive Neuroscience, vol. 3, pp. 71-86, 1991.
- [TS06] X. Tana, C. Songcan. "Face recognition from a single image per person: A survey". Pattern Recognition, vol. 39, n° 9, pp. 1725-1745, 2006.
- [VAET97] D. Valentin, H. Abdi, B. Edelman, A.J. O'Toole, "Principal component and neural network analyses of face images: What can be generalized in gender classification?", Journal of Mathematical Psychology, Vol.41, Issue 4, pp.398-412, 1997.

- [Vap95] V.N. Vapnik, "**The Nature of Statistical Learning Theory**", Springer-Verlag, 188 pages, New York, 1995.
- [Vap98] V. N. Vapnik, "**Statistical Learning Theory**", Wiley-Interscience, 736 pages, 1998.
- [Vap00] V. N. Vapnik, "**The Nature of Statistical Learning Theory**", 2nd ed., Springer-Verlag New York Inc., 314 pages, 2000.
- [Vis05] M. Visani, "Vers de nouvelles approches discriminantes pour la reconnaissance automatique de visages", Thèse de Doctorat, Institut National des Sciences Appliquées de Lyon, France, 2005.
- [Vis.fr] VISIONIC, système de vision et contrôle sans contact, <http://www.visionic.fr>.
- [VJ01] P. Viola, M. Jones, "*Rapid Object Detection using a Boosted Cascade of Simple Features*", IEEE Computer Vision and Pattern Recognition, vol. 1, pp. 511-518, 2001.
- [WAH03] T. Wang, H. Ai, G. Huang, "*A Two-Stage Approach to Automatic Face Alignment*", Proc. Third Int'l Symp. Multispectral Image Processing and Pattern Recognition, H. Lu and T. Zhang, eds., pp. 558-563, 2003.
- [WBMR08] J. Wu, S. C. Brubaker, M. D. Mullin, J. M. Rehg, "*Fast Asymmetric Learning for Cascade Face Detection*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 3, pp. 369-382, 2008.
- [Wec06] H. Wechsler, "*Reliable Face Recognition Methods, System Design Implementation and Evaluation*", Springer, 2006.
- [WFKM95] L. Wiskott, J.M. Fellous, N. Krüger, C. Malsburg, "*Face recognition and gender determination*", In Proc. of the Int. Workshop on Automatic Face- and Gesture-Recognition, pp. 92-97, Zurich, Suisse, 1995.
- [Wik.org] Le protocole de communication X10, [http://fr.wikipedia.org/wiki/X10_\(informatique\)](http://fr.wikipedia.org/wiki/X10_(informatique)).
- [Wis97] L. Wiskott, J.-M. Fellous, N. Krüger, C. von der Malsburg, "*Face recognition by elastic bunch graph matching*". IEEE Trans. Patt. Anal. Mach. Intell. 19, 775-779, 1997.
- [WPB⁺98] H. Wechsler, P.J. Phillips, V. Bruce, F.F. Soulie, T.S. Huang, "*Face Recognition : from Theory to Applications*", pp.51-72, Springer-Verlag, 1998.
- [WS08] J. Wu, W. A. P. Smith, E. R. Hancock, "*Gender Classification based on Facial Surface Normals*", the 19th International Conference on Pattern Recognition, Florida, USA, 2008.
- [WY06] Y. Wu, T. Yu, "*A Field Model for Human Detection and Tracking*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 5, pp. 753-765, May, 2006.
- [XLS08] Z. Xu, L. Lu, P. Shi, "*A Hybrid Approach to Gender Classification from Face Images*", the 19th International Conference on Pattern Recognition, Florida, USA, pp. 1-4, 2008.
- [YAK01] M. Yang, N. Ahuja, D. Kriegman, "*Face Detection Using Multimodal Density Models*", Computer Vision and Image Understanding, pages 264-284, 2001.
- [Yan02] M.H. Yang, "*Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods*", Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp.215-220, 2002.

- [YKA02] M. Yang, D. Kriegman, N. Ahuja, "*Detecting Faces in Images : A Survey*", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 24, n° 1, pp. 34-58, 2002.
- [YLA06] Z. Yang, M. Li, H. Ai, "*An Experimental Study on Automatic Face Gender Classification*", Proc. 18th IEEE Int'l Conf. Pattern Recognition, vol. 3, pp. 1099-1102, 2006.
- [ZCK98] W. Zhao, R. Chellapa, A. Krishnaswamy, "*Discriminant analysis of principal components for face recognition*", Proceedings International Conference on Automatic Face and Gesture Recognition, pp.336-341, 1998.
- [Zer34] F. Zernike, "*Diffraction theory of the cut procedure and its improved form, the phase contrast method*". Physica, 1: pp. 689-704, 1934.
- [Zha99] W. Zhao, "Robust Image Based 3D Face Recognition", PhD thesis, University of Maryland, 1999.
- [Zha03] W.Zhao, R. Chellapa, P. J. Phillips, A. Rosenfeld, "*Face Recognition: A Literature survey*", ACM Computing Surveys, Vol. 35, No. 4, pp. 399-458, 2003.
- [ZRL96] T. Zhang, R. Ramakrishnan, M. Livny, "*BIRCH: An Efficient Data Clustering Method for Very Large Databases*", Int. Conf. Management of Data (SIGMOD'96), pp. 103-114, Montreal, Canada, June 1996.
- [ZRL97] T. Zhang, R. Ramakrishnan, M. Livny, "*BIRCH: A New Data Clustering Algorithm and Its Applications*", Data Mining and Knowledge Discovery, vol. 1, n° 2, pp. 141-182, 1997.

ANNEXES

Annexe A

Développement d'une base de données biométriques : le projet IV²

Le projet IV² a pour but de créer une base de données de référence contenant des séquences d'images d'iris et du visage, capturées par plusieurs caméras synchrones, permettant ainsi de reconstituer l'information 3D du visage par stéréovision. Il a pour but également de créer les conditions d'une évaluation à l'échelle internationale de différents systèmes de reconnaissance biométrique liés à l'information du visage, de l'iris et de la voix, dans des milieux semi-contraints, à savoir :

- des systèmes d'identification par l'iris ;
- des systèmes d'identification par le visage 2D et 3D ;
- des systèmes de détection des yeux dans les images 2D ;
- des systèmes de reconstruction 3D du visage ;
- des systèmes d'identification de visages parlants ;

Le projet IV² s'est articulé autour de différentes phases. Nous avons commencé par la conception et la réalisation de la cabine d'acquisition. L'enregistrement des données personnelles dans le cadre du projet IV² a nécessité l'accord de la Commission Nationale de l'informatique et des Libertés, une procédure administrative qui a duré une année et qui a retardé le développement du projet. Notre demande de constitution de la base de données multimodale et sa dissémination auprès des chercheurs français et étrangers a reçu un avis favorable, pour trois ans, le 18 janvier 2007. Nous avons défini, à l'issue de plusieurs réunions entre les partenaires du projet, les protocoles d'acquisition, les données enregistrées et les évaluations des méthodes de reconnaissances sur celles-ci.

A.1 Conception et réalisation de la cabine d'acquisition

Le projet IV² ayant pour but de constituer une base de données de différentes modalités biométriques, la première étape consistait donc à concevoir et à réaliser un système d'acquisition permettant de satisfaire à ces exigences. En effet si les capteurs étaient disponibles sur le marché, il fallait réaliser une structure souple, fiable, transportable, permettant d'effectuer en un temps restreint les différentes acquisitions et de stocker les données ainsi acquises. Nous sommes donc arrivés à un projet de « photomaton » biométrique.

A.1.1 Cahier des charges de la cabine

Avant toute chose, il a fallu se mettre d'accord avec les divers partenaires du projet, afin de définir les spécificités de la cabine, par exemple la taille adéquate de celle-ci, les conditions

d'éclairages ou de prises de vues. Deux réunions ont été nécessaires afin de pouvoir arriver à un projet validé par l'ensemble des partenaires du consortium.

Pour acquérir les données sélectionnées, le projet propose le développement d'une cabine multi-capteurs, selon la description suivante : quatre caméras vidéo pour la stéréovision des images de visage de face et de profil (visages 2D & 3D), une caméra infrarouge pour l'acquisition des images d'iris, un scanner 3D (visages 3D), un caméscope numérique et une webcam pour les acquisitions des visages parlants synchronisés à l'acquisition de la voix. Par ailleurs ce «photomaton» biométrique doit pouvoir être déplacé dans différents lieux pour réaliser les acquisitions. Sa structure doit donc être légère et aisément démontable. Son encombrement au sol doit être restreint.

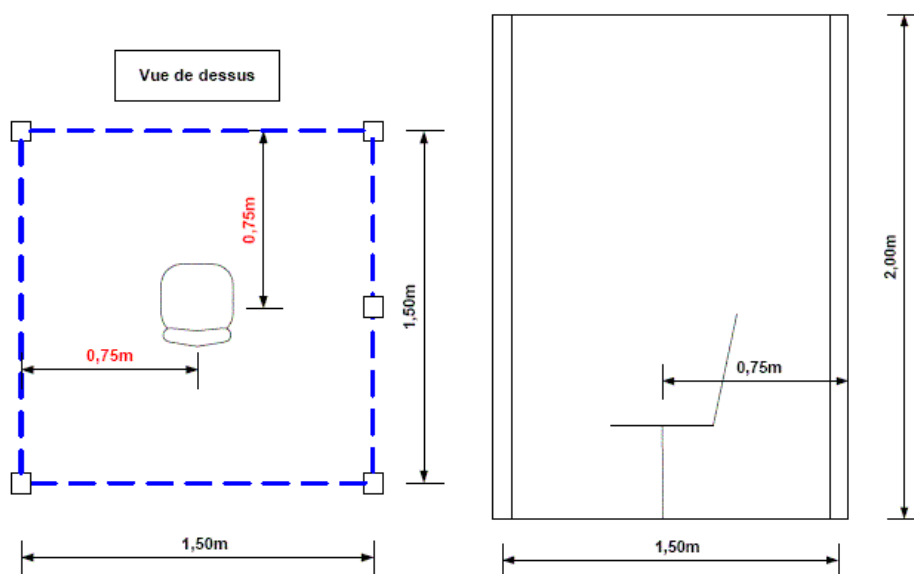


Figure A.1 : Vue de dessus, et vue de profil de la cabine.

A.1.2 Structure de la cabine

La cabine doit avoir une taille correcte pour que tous les éléments puissent être installés à l'intérieur, et surtout pour qu'elle soit transportable facilement. Nous avons donc retenu une taille de 1,50m de longueur et de largeur, et 2m de hauteur (voir figure 4.1). Cette taille est adéquate, car elle offre les bonnes distances pour les prises de vues de chaque capteur.

Pour assurer la mobilité de la cabine et en faciliter le montage/démontage, et le transport, celle-ci a été réalisée avec une structure en profilé d'aluminium (figure 4.2). Ce type de structure est très utilisé en industrie, car elle est très robuste, et surtout très légère. La légèreté de ce produit est un atout pour permettre un transport aisé de la structure. Le schéma de la structure est donné sur la figure 4.2. L'entrée de la cabine est situé sur le côté. Pour la stabilité de la cabine au sol, et afin de pouvoir la régler correctement, un socle en fonte d'aluminium est monté à chaque pied.

Afin d'avoir des conditions d'éclairages optimales, 4 halogènes sont disposés sur les traverses supérieures, orientés vers le sol et 2 spots sont installés en face de la personne, sur les montants de la cabine afin de limiter les ombres formées par le nez. Les halogènes sont gérés par des télévariateurs contrôlés par l'ordinateur gérant les acquisitions afin de pouvoir créer des conditions d'éclairage variables à des instants précis du protocole. Ce système, basé sur le

protocole de communication X10 [Wik.org] est utilisé dans des applications domotique. Un logiciel permet de réaliser les commandes en se basant sur des évènements particuliers. Dans notre cas, nous avons utilisé l'horloge interne du PC pour rythmer la séquence de variation des éclairages que nous détaillons dans le paragraphe 4.2.2.3. Afin d'empêcher la lumière du jour de venir gêner les prises de vues, des rideaux opaques sont fixés autour de la structure de la cabine. Ces rideaux sont de couleur gris à 17% coté intérieur. Cette couleur a été retenue car elle est très utilisée dans le milieu photographique pour les fonds d'images, et elle permet de bien distinguer une personne du fond de la cabine. Pour des raisons de sécurité les rideaux sont réalisés avec une matière ininflammable. La figure 4.3 montre la cabine opérationnelle telle que nous l'avons montée dans le cadre du projet IV².



Figure A.2 : Structure de la cabine et disposition de l'éclairage.



Figure A.3 : « Photomaton » biométrique du projet IV².

A.1.3 Les capteurs

Une fois la structure validée par l'ensemble des partenaires, il a fallu choisir les différents capteurs permettant d'acquérir les données biométriques retenues.

Visages 2D et 3D par stéréovision. L'acquisition en 2D de face et de profil a été réalisée à l'aide de deux bancs vidéo à tête stéréoscopique (voir figure 4.4). L'acquisition a une durée de 40 secondes par personnes. Les caméras faisant cette acquisition se devaient d'être de très bonne qualité, avec des objectifs à faible distorsion. La solution retenue pour les caméras, est la suivante : les caméras couleurs sont numériques, et possèdent une résolution de 640 par 480. Il faut savoir que celles-ci sont toutes synchronisées entre elles par la carte d'acquisition. Cette synchronisation est indispensable pour permettre d'effectuer ultérieurement une reconstruction 3D de qualité.

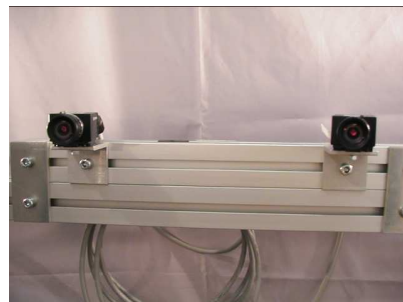


Figure A.4 : Une caméra et un des bancs stéréoscopiques installés dans la cabine.

Pour permettre de passer du 2D au 3D à partir des caméras équipant les bancs stéréoscopiques, nous avons mis en oeuvre une procédure de calibration basée sur la méthode de Toscani-Faugeras [MH95, FT86]. Les paramètres des matrices intrinsèques et extrinsèques des caméras ainsi que ceux des matrices de transformation sont inclus dans chaque répertoire de la base de données.

Visages 3D par scanner. L'acquisition en 3 dimensions du visage sera réalisée à l'aide d'un scanner 3D Minolta (figure 4.5), acquis par le partenaire LIRIS-ECL. Ce scanner, branché sur une prise USB, permet de réaliser conjointement une image de profondeur et une image de texture du visage d'une personne. Ces prises de vues sont réalisées à l'aide d'un laser de classe I, non dangereux pour les yeux et le corps humain. Les logiciels d'acquisition et de reconstruction 3D sont fournis par le constructeur. La reconstruction 3D s'effectue en différé à partir des 3 vues de face, de profil gauche et de profil droit.



Figure A.5 : Le scanner permettant l'acquisition des visages 3D.

Visages parlants. L'acquisition du visage parlant se fait à l'aide de deux capteurs : un caméscope numérique pour obtenir une séquence vidéo de bonne qualité et une webcam permettant d'obtenir une séquence en mode dégradé (figure 4.6). Le caméscope numérique retenu enregistre les séquences sur des micros cassettes au format DV. Lors des phases d'acquisition, il a été donc nécessaire de transférer manuellement les données depuis cette cassette dans la mémoire du PC de commande.



Figure A.6 : Le caméscope et la webcam (au milieu du banc stéréoscopique de face) permettant d'enregistrer la modalité visage parlant.

Iris. Le mode d'acquisition retenu pour l'iris a été celui passant par une caméra infrarouge. Celle-ci, de fabrication coréenne, a été fournie par les partenaires du GET-INT (figure 4.7). Les données transitent par une prise USB. Un logiciel d'acquisition dédié est disponible avec la caméra.

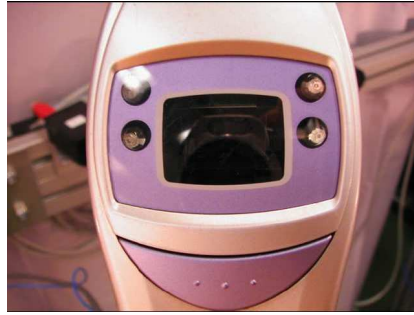


Figure A.7 : La caméra infrarouge permettant l'acquisition de l'iris.

A.1.4 Sauvegarde des données

La base de données IV² est de grande taille comme le montre le tableau 4.1. Le stockage sur le PC de commande ne représente qu'une solution intermédiaire, les données étant régulièrement transférées vers un disque dur externe de grande capacité (1To). Deux disques durs de cette taille sont utilisés pour sauvegarder la base de données et pouvoir travailler dessus sans risque de perdre les données acquises.

Modalité	Taille du fichier / Personne
Visages 2D (2 bases stéréoscopiques)	2000 Mo
Visage 3D (Scanner Minolta)	11 Mo
Visage parlant (Caméscope Numérique)	200 Mo
Visage parlant (Webcam)	15 Mo
Iris (Caméra Infrarouge)	12 Mo
Taille moyenne pour 1 personne	2,2 Giga Octets

Tableau A.1 : Taille moyenne des données pour une personne

A.2 Les protocoles d'acquisition

Lorsqu'une personne se présente dans la cabine, les objectifs du projet et les protocoles d'acquisition lui sont proposés. Les personnes volontaires remplissent un formulaire comportant des renseignements sur leur identité ainsi que des renseignements sur leur couleur de peau, leur couleur de cheveux, la présence d'individus de leur famille dans la base,... Ces informations sont ensuite saisies et stockées dans une base de données. A chaque acquisition, un identifiant est automatiquement calculé par le système informatique. C'est cet identifiant qui servira de nom au répertoire contenant les données biométriques multimodales de la personne. Pour des raisons de confidentialité, le fichier contenant les données personnelles est conservé à part de la base de données IV². Mais les informations génériques liées à la couleur de la peau, couleur des cheveux et des yeux, signes particuliers et liens de parenté sont des informations qui peuvent être utiles aux utilisateurs de la base de données IV². En effet ce type d'informations peut permettre de sélectionner une classe d'individus particulier sur lesquels on veut évaluer un traitement : classification suivant l'âge, le genre, l'origine ... Cette partie des informations est donc extraite de manière anonyme et va venir abonder un répertoire dit de «méta données» qui est créé dans le répertoire racine portant l'identifiant de la personne concernée par ces informations. La séquence d'acquisition des données commence ensuite dans l'ordre ci-dessous :

- Visage 3D - Scanner 3D
- Visage parlant - Webcam/caméscope
- Visage 2D/3D - Bases stéréoscopiques et caméscope
- Iris – Caméra infrarouge

A.2.1 Visage 3D – Scanner 3D

Le premier objectif de la prise d'information avec le scanner 3D est de pouvoir disposer d'un modèle 3D complet de la face. Ceci se fait à partir des trois premières vues, de face, de $\frac{3}{4}$ droit et de $\frac{3}{4}$ gauche, acquises du visage neutre. Les autres vues acquises permettent de comparer des cartes de profondeur, ou des couples cartes de profondeur/images de texture, à ce modèle 3D reconstruit. Pour cela une première capture de la vue de face avec les yeux fermés et sans expression est réalisée. Ensuite, l'objectif est de mesurer la variabilité des algorithmes de reconnaissance face aux problèmes des expressions. Trois captures, pendant lesquelles on demande au volontaire de simuler une expression particulière, sont alors réalisées dans ce but : 1 face « sourire », 1 face « surprise », 1 face « dégoût ». Pour finir, on réalise une capture de la face neutre sans éclairage. L'objectif étant de mesurer la reconnaissance avec pour seule information la carte de profondeur, l'image de texture étant trop sombre pour permettre alors une quelconque utilisation. La figure 4.8 présente quelques exemples des vues disponibles en cours d'acquisition.



Figure A.8 : Acquisition d'une face par le scanner 3D

A.2.2 Visage parlant - Caméscope/Webcam

Le principe sous entendu par la modalité «visage parlant» consiste à combiner les informations de la voix à celles liées au visage. Le principe le plus courant étant de fusionner les scores issus de deux processus disjoints de reconnaissance. Mais, pour rendre des systèmes d'authentification plus résistant aux attaques, il est parfois envisagé d'utiliser le couplage existant entre le mouvement des lèvres et les phrases prononcées.

Les enregistrements par le caméscope et la webcam débutent en parallèle. Dans une ambiance convenablement éclairée, la personne lit 15 phrases proposées qui balayent les différents phonèmes de la langue française. A l'issue de la lecture, qui dure de 1 minute à 1 minute 30, selon les personnes, il est demandé au volontaire de regarder à sa droite de manière à présenter son profil gauche aux deux capteurs, puis de regarder à sa gauche, de manière à présenter son profil droit et pour finir, de regarder le plafond et le plancher. Ainsi obtient on des variations de pose dans ces deux séquences vidéo. A noter que la webcam est coupée à la fin de la séquence que nous venons de décrire, alors que le caméscope continue à enregistrer pendant toute la séquence suivante qui concerne les bases stéréoscopiques et qui va comporter des variations de pose, d'expression et d'éclairage.

L'intérêt d'utiliser le caméscope et la webcam est illustré sur la figure 4.9. En effet le caméscope procure une image de bonne qualité, alors que la webcam propose une vision dégradée de la même scène. L'objectif est donc de tester des algorithmes dans ces conditions différentes.

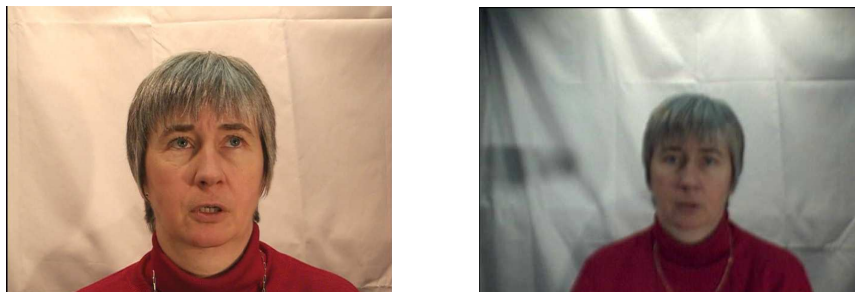


Figure A.9 : Images provenant des séquences prises par le caméscope numérique et par la webcam.

A.2.3 Visage 2D/3D – Bases stéréoscopiques et caméscope

L'objectif des données acquises à partir des bases stéréoscopiques est de disposer de quatre images simultanées d'une même personne (figures 4.10 et 4.11). Deux de face et deux de profil. Cela donne ainsi une information redondante en 2D. Le fait que les bases soient calibrées procure aussi la possibilité d'opérer une reconstruction en 3D, de face et de profil de cette même personne. Cela permet d'envisager plusieurs processus de reconnaissance à partir d'une ou plusieurs images 2D, mais aussi en couplant des informations 2D de face et de profil ou en couplant des informations 2D et 3D. La voie est donc ouverte à de nombreuses évaluations. La séquence d'acquisition, qui dure une quarantaine de secondes, contient des variations de poses, d'expressions et d'éclairage. Les consignes données au volontaire sont les suivantes :

- Début d'acquisition en lumière maximum, la personne étant de face.
- Puis la personne doit réaliser un trajet visuel qui part du centre et qui va balayer différentes icônes positionnées de gauche à droite avec pour consigne d'imiter les expressions des icônes (peur, dégoût, sourire). Ce trajet dure une dizaine de secondes, la lumière étant au maximum au début et commence une décroissance à travers trois ou quatre paliers qui durera dix secondes. La personne étant de face sur la fin de cette décroissance.
- 10 secondes avec bascule des éclairages sur les halogènes latéraux, 5 s. droite et 5 s. gauche, personne de face.
- 15 secondes finales, plein éclairage, personne de face mais qui est libre de ses mouvements ce qui entraîne une faible variation de pose ou d'expressions.

Remarquons que durant l'ensemble de cette séquence, le caméscope numérique est toujours en mode enregistrement. Cela nous donne encore une fois la possibilité, en scindant la séquence caméscope, de comparer des acquisitions issues de deux capteurs de qualités différentes.



Figure A.10 : Les deux vues délivrée par les caméras de la base stéréoscopique « face »



Figure A.11 : Les deux vues délivrée par les caméras de la base stéréoscopique « profil »

A.2.4 Iris

L'acquisition des iris au moyen de la caméra infrarouge (figure 4.12) obéit au protocole suivant. Pour chaque personne, 2 sessions sont réalisées pour chacun des deux yeux (gauche : G1 et G2, et droit : D1 et D2).

- La première session (G1 et D1) consiste en 10 images qualifiées prises sans lunettes, si le donateur porte des lunettes.
- La deuxième session (G2 et D2) consiste en 10 images qualifiées prises avec lunettes, pour les donateurs en portant, (sans lunettes pour les autres).

Remarquons que pour chaque session et chaque acquisition (gauche ou droite) le logiciel accompagnant la caméra iris, crée deux répertoires : un contenant les images qualifiées (nombre d'images compris en 10 et 20) et l'ensemble des images acquises, qui pour un grand nombre d'entre elles ne sont pas jugées de qualité suffisante par le capteur. Il est donc encore une fois possible d'envisager la reconnaissance en mode dégradé.



Figure A.12 : Photo d'un iris prise par la caméra infrarouge

Annexe B

Centre, rayon et diamètre d'une classe

Etant donné N points de données d -dimensionnels dans une classe C , $\{\vec{X}_i\}$ où $i=1,2,\dots,N$, le centre \vec{X}_0 , le rayon R et le diamètre D de la classe sont définis comme suit :

$$\vec{X}_0 = \frac{\sum_{i=1}^N \vec{X}_i}{N} \quad (\text{A.1})$$

$$R = \left(\frac{\sum_{i=1}^N (\vec{X}_i - \vec{X}_0)^2}{N} \right)^{1/2} \quad (\text{A.2})$$

$$D = \left(\frac{\sum_{i=1}^N \sum_{j=1}^N (\vec{X}_i - \vec{X}_j)^2}{N(N-1)} \right)^{1/2} \quad (\text{A.3})$$

R est la distance moyenne entre les points de données de la classe C et le centre. D est la distance paire moyenne au sein de la classe C . Elles sont deux mesures alternatives de la compacité de la classe autour du centre. Ensuite entre deux classes, on définit cinq distances alternatives afin de mesurer leur proximité. Soit les centres de deux classes : \vec{X}_{0_1} et \vec{X}_{0_2} , la distance euclidienne D_0 et la distance Manhattan D_1 entre les deux centres relatifs aux deux classes sont définies comme suit :

$$D_0 = \left((\vec{X}_{0_1} - \vec{X}_{0_2})^2 \right)^{1/2} \quad (\text{A.4})$$

$$D_1 = \left| \vec{X}_{0_1} - \vec{X}_{0_2} \right| = \sum_{i=1}^d \left| \vec{X}_{0_1}^{(i)} - \vec{X}_{0_2}^{(i)} \right| \quad (\text{A.5})$$

Soit N_1 points de données d -dimensionnels dans une classe C_1 : $\{\vec{X}_i\}$ où $i=1,2,\dots,N_1$, et N_2 points de données dans une autre classe C_2 : $\{\vec{X}_j\}$ où $j=N_1+1, N_1+2,\dots, N_1+N_2$, la distance

inter-classes moyenne D_2 , la distance intra-classes moyenne D_3 et la distance de l'augmentation des variances des deux classes D_4 sont définies ainsi :

$$D_2 = \left(\frac{\sum_{i=1}^{N_1} \sum_{j=N_1+1}^{N_1+N_2} (\overline{X}_i - \overline{X}_j)^2}{N_1 N_2} \right)^{1/2} \quad (\text{A.6})$$

$$D_3 = \left(\frac{\sum_{i=1}^{N_1+N_2} \sum_{j=1}^{N_1+N_2} (\overline{X}_i - \overline{X}_j)^2}{(N_1 + N_2)(N_1 + N_2 - 1)} \right)^{1/2} \quad (\text{A.7})$$

$$D_4 = \left(\sum_{k=1}^{N_1+N_2} \left(\overline{X}_k - \frac{\sum_{i=1}^{N_1+N_2} \overline{X}_k}{N_1 + N_2} \right)^2 - \sum_{i=1}^{N_1} \left(\overline{X}_i - \frac{\sum_{i=1}^{N_1} \overline{X}_i}{N_1 + N_2} \right)^2 - \sum_{j=N_1+1}^{N_1+N_2} \left(\overline{X}_j - \frac{\sum_{i=N_1+1}^{N_1+N_2} \overline{X}_i}{N_2} \right)^2 \right)^{1/2} \quad (\text{A.8})$$

On considère \overline{X}_0 , R et D comme les propriétés d'une seule classe, et D_0, D_1, D_2, D_3 et D_4 comme les propriétés de deux classes.

Ce qui suit c'est les deux mesures alternatives de la qualité de la classification : Rayon moyen au carré pondéré Q_1 (ou \overline{R}), et Diamètre moyen au carré pondéré Q_2 (ou \overline{D}).

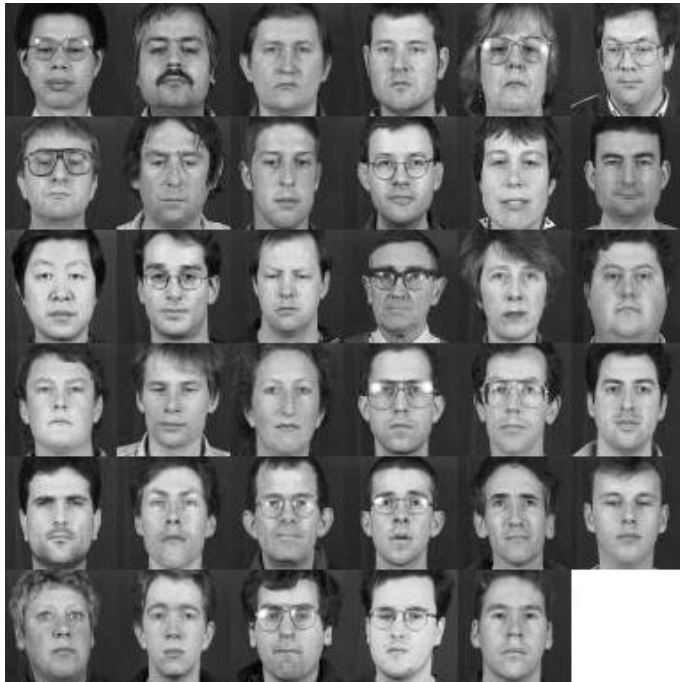
$$Q_1 = \frac{\sum_{i=1}^K n_i R_i^2}{\sum_{i=1}^K n_i} \quad (\text{A.9})$$

$$Q_2 = \frac{\sum_{i=1}^K n_i (n_i - 1) D_i^2}{\sum_{i=1}^K n_i (n_i - 1)} \quad (\text{A.10})$$

Annexe C

Classes d'individus

Les six dernières classes de l'exemple de partitionnement établi et illustré au chapitre 5.



Classe n°5



Classe n°6



Classe n°7



Classe n°8



Classe n°9



Classe n°10

Annexe D

Algorithme d'insertion d'une entrée dans un CF-tree

Rappel de l'algorithme de BIRCH

L'algorithme d'insertion d'une entrée Ent (un unique point de données ou une sous classe) dans un $CF-tree$ se présente comme suit :

1. **Identifier la feuille (leaf) appropriée** : en commençant par la racine, descendre le $CF-tree$ récursivement en choisissant le nœud-fils le plus proche. Dans notre développement, nous avons choisi la métrique D_0 défini dans [ZRL97] comme étant la distance Euclidienne des centroïdes.
2. **Modifier la feuille** : en atteignant un nœud feuille, trouver l'entrée feuille la plus proche, soit L_i , ensuite vérifier si L_i peut absorber Ent sans violer la condition du seuil. Ce qui veut dire que la classe fusionnée - Ent et L_i - doit satisfaire la condition du seuil. Si c'est le cas, mettre à jour l'entrée CF par L_i . A noter que l'entrée CF de la nouvelle classe peut être calculée à partir des entrées CF pour L_i et Ent . Sinon, ajouter à la feuille une nouvelle entrée Ent . Si l'on dispose d'assez d'espace sur la feuille pour y loger cette nouvelle entrée, c'est tant mieux, sinon on doit diviser le nœud feuille. Cette division se fait en sélectionnant comme centres de nœuds dérivants la paire d'entrées la plus distante et redistribuant les entrées restantes en se basant sur les distances minimales.
3. **Modifier le chemin vers la feuille** : après avoir inséré Ent dans la feuille, mettre à jour l'information CF pour chaque nœud non feuille sur son chemin vers la feuille. L'absence de division implique la mise à jour des nœuds existants par l'injection de Ent . La division d'une feuille nous oblige à insérer une nouvelle entrée non feuille dans le nœud parent afin de décrire la nouvelle feuille créée. Si le parent avait de l'espace pour cette entrée, à tous les niveaux supérieurs, nous n'avons qu'à mettre à jour les entrées CF pour refléter l'addition de Ent . Cependant, nous pouvons nous trouver obligés de diviser le parent, et ainsi de suite jusqu'à la racine. Si la racine est divisée, la hauteur de l'arbre augmente de un.
4. **Le raffinement d'une fusion** : Si on est en présence d'un arbre oblique des données, la qualité de la classification et la bonne utilisation de l'espace disponible vont certainement dégrader. Une simple étape supplémentaire de fusion aide souvent à résoudre ce problème : supposons qu'on effectue une division de feuilles et que la propagation de cette division s'arrête à un nœud non feuille quelconque N_j , c'est-à-dire N_j peut héberger l'entrée supplémentaire résultante de la division, nous pouvons alors étudier le nœud N_j pour trouver les deux entrées les plus proches. S'il s'avère qu'elles ne sont pas la paire qui correspond à la division, nous essayerons de les fusionner ainsi que leurs nœuds fils correspondant. S'il existe plus d'entrées dans les deux nœuds que le facteur bifurquant B , nous divisons le résultat de la fusion. Pendant cette re-division, en cas où l'un des deux noyaux attire assez d'entrées fusionnées, nous mettons le reste des entrées avec l'autre

noyau. En résumé, si les entrées fusionnées, tiennent dans un seul nœud, nous libèrerons un espace pour une entrée supplémentaire dans le nœud N_j . Ceci augmente l'utilisation de l'espace en remettant à plus tard les divisions futures et améliore la distribution des entrées dans les fils les plus proches.

Ces étapes d'insertion permettent d'ajuster d'une manière dynamique l'arbre CF pour diminuer sa sensibilité à l'ordre de l'alimentation des données.

Annexe E

Implémentation de BIRCH

Les classes principales du programme sont décrites dans le tableau ci-dessous :

Nom de la classe	Descriptions
CCFTree	C'est l'arbre de CF. Il se compose d'un CCFNode qui joue le rôle de Racine et un ensemble de nœuds CCFNode. CCFTree supporte les opérations de construction de l'arbre (insertion, reconstruction) et d'exportation des résultats de la classification (exporter les feuilles de l'arbre qui sont des sous classes et la structure hiérarchique de l'arbre de CF sous forme de XML)
CCFNode	C'est le nœud de CF : il y a deux types de nœuds : non feuille et feuille dont la différence principale est le nombre maximale d'entrées qu'ils contiennent (taille du nœud). Un nœud possède d'une liste d'entrées CCFEntryList.
CCFEntryList	C'est la liste d'entrées d'un nœud. Elle sert à gérer les opérations d'insertion, de suppression des entrées d'un nœud.
CCFEntry	C'est une entrée dans un nœud qui contient un vecteur de CF. Elle est une sous classe créée lors de la classification. Un point de données peut être considéré comme une entrée libre. Elle peut être insérée à un nœud, se déplacer d'un nœud à un autre, être absorbée par une autre entrée.
CCFVector	C'est le vecteur de CF. Il est une collection d'informations concernant une sous classe. Il supporte les opérations d'addition, de soustraction, le calcul de compacité, la distance entre lui et un autre vecteur de CF. Cette classe est conçue virtuelle. Un vecteur réel peut contenir les informations concrètes concernant la sous classe (par exemple CCFNumVector pour les données numériques ou un autre pour les données catégorie, mélangées,...)

Classe CCFTree

Pour construire un arbre de CFs, il faut fournir les informations suivantes (arguments d'entrée) :

Argument	Type	Description
treeSize	Entier	La taille de l'arbre, mesurée par le <i>nombre de nœud maximum</i> que l'arbre peut contenir.
nodeSize	Entier	La taille d'un nœud dans l'arbre, mesurée par le <i>nombre d'entrées maximum</i> qu'une feuille peut contenir.
leafSize	Entier	La taille d'une feuille dans l'arbre, mesurée par le <i>nombre d'entrées maximum</i> qu'une feuille peut contenir.
dataDim	Entier	<i>Le nombre de dimensions</i> des données traitées.
outlierThres	Entier	Le seul pour déterminer des aberrants. C'est le <i>nombre maximum de point</i> de données d'une entrée feuille afin de décider si une entrée feuille est considérée comme une sous classe ou un aberrant. C'est-à-dire si le nombre de points contenus dans une entrée feuille est moins que cette valeur, alors cette entrée est un aberrant.
maxOutSaved	Entier	L'espace de disque réservé à stocker temporairement les aberrants potentiels, mesuré par le <i>nombre d'entrées feuilles maximum</i> qui sont considérées comme des aberrants qu'on veut sauvegarder sur le disque.
maxDelaySaved	Entier	L'espace de disque réservé à stocker temporairement les points retardés lors de leur insertion dans l'arbre, mesuré par le <i>nombre de points maximum</i> qu'on veut stocker sur le disque en attendant la reconstruction de l'arbre.
Threshold	Double	Le seuil d'absorption des entrées de feuilles de l'arbre. Il vaut mieux mettre cette valeur à zéro au début de la construction de l'arbre afin d'obtenir un arbre plus compact que possible.

En utilisant ces informations d'entrée, un arbre de CFs avec le nœud de racine sera créé et prêt à accueillir les points qui arrivent pour former un arbre de CFs dont les nœuds sont des

sous-classes contenant les points de données qui sont les plus similaires. Le degré de raffinement des sous-classes (la compacité ou le diamètre) est réduit de bas en haut.

Les opérations principales supportées dans la classe CCFTree.

Nom	Description
insertEntry	Insérer une entrée dans l'arbre en visitant tout d'abord le nœud de racine et descendant dans l'arbre jusqu'à une feuille appropriée.
getNewThreshold	Calculer le nouveau seuil d'absorption pour la reconstruction de l'arbre.
reinsertDelays	Réinsérer les points retardés stockés sur le disque après avoir reconstruit l'arbre.
reinsertOutliers	Réinsérer les classes considérées comme des aberrants potentiels stockés sur le disque après avoir reconstruit l'arbre.
saveOutlier	Sauvegarder une sous classe considérée comme un aberrant potentiel sur le disque pendant la reconstruction de l'arbre.
saveDelay	Sauvegarder un point qui ne peut pas être inséré dans l'arbre sans augmenter la taille de l'arbre sur le disque.
Rebuild	Reconstruire l'arbre avec le nouveau seuil d'absorption.

Pour insérer un point de données dans l'arbre, tout d'abord, ce point est représenté comme une entrée.

L'arbre de CFs insère une entrée en commençant au nœud de racine, en descendant récursivement pour trouver un nœud de feuille approprié. Cette entrée est soit insérée à un nœud de feuille comme une nouvelle entrée, soit absorbée par une entrée existante dans ce nœud. Si cette opération ne fait pas dépasser le nombre de nœud de l'arbre, alors un autre point peut être inséré. Au contraire, on fait retarder ce point en vérifiant si l'espace réservé aux points retardés est encore disponible. S'il l'est, le point est sauvegardé sur le disque. S'il ne l'est pas, c'est-à-dire que n'importe quel point qui arrive n'est ni inséré dans l'arbre car l'espace de l'arbre peut être dépassé sa capacité disponible s'il est inséré, ni sauvegardé sur le disque comme un point retardé qui attend une nouvelle chance d'être inséré à l'arbre, alors que il faut reconstruire l'arbre en augmentant le seuil d'absorption des entrées. Après avoir reconstruit l'arbre, on réinsère les points retardés stockés au disque au nouvel arbre de CFs, si un point retardé n'est pas réinséré, il est encore considéré comme un point retardé et il reste au disque.

Pour calculer le nouveau seuil, le programme extrait tous les nœuds de feuille de l'arbre actuel. A chaque nœud de feuille, pour chaque entrée, on cherche l'entrée la plus proche dans le même nœud. Pour toutes les paires d'entrée les plus proches, on calcule la somme de leurs distances. Si un nœud contient une seule entrée, alors on cherche l'entrée la plus proche d'elle dans tous les nœuds de feuille de l'arbre en profitant des liens précédents et suivants de chaque nœud de feuille. L'augmentation du seuil est la moyenne de la somme totale de distances de tous les pairs d'entrées les plus proches dans tous les nœuds de feuille. En utilisant ce nouveau seuil, on espère que la moitié de nœuds de feuille est réduite.

Après avoir terminé la construction d'un arbre de CFs, on obtient les informations suivantes (sorties) :

Nom	Type	Description
Cftree.xml	Fichier XML	L'image de l'arbre de CFs obtenu est représentée sous la forme d'un fichier XML qui exprime très bien la structure hiérarchique d'arbre de CFs qu'on obtient à la fin de la construction de l'arbre.
Clusters.txt	Fichier Textuel	Les sous classes produites de la construction de l'arbre.
Outliers.txt	Fichier textuel	Les sous classes considérées comme des vrais aberrants qu'on obtient à la fin de la construction de l'arbre.

Classe CCFNode

Pour construire un nœud, il faut fournir les informations suivantes :

Argument	Type	Description
Size	Entier	La taille de nœud, mesurée par le nombre d'entrées maximum qu'il peut contenir.
threshold	Double	Le seuil d'absorption des entrées contenues dans le nœud.

En utilisant ces informations, on construit un nœud avec une taille et un seuil d'absorption précisé. Normalement, seuls les nœuds de feuille sont liés par deux liens : un vers le nœud précédé, un vers le nœud suivant. En plus, la différence entre deux types de nœuds (non feuille et feuille) est également la taille spécifiée par l'arbre de CFs lors de l'insertion d'un point à l'arbre. Une liste d'entrée vide à taille du nœud est créée pour stocker les sous-classes

groupées dans le nœud. Le parent d'un nœud est une entrée. Initialement, le parent d'un nœud est nul.

Les opérations principales de CFFNode :

Nom	Description
Insert	Insérer une entrée dans l'arbre. A un nœud, il faut d'abord chercher l'entrée la plus proche parmi toutes les entrées situées dans le nœud. Si l'entrée trouvée est une entrée feuille, alors on décide si cette entrée est absorbée ou insérée au nœud. Au contraire, on descend au nœud fils de cette entrée.
getClosest	Chercher l'entrée la plus proche d'une entrée à insérer dans un nœud en calculant les distances entre l'entrée à insérer avec toutes les entrées ou une partie d'entrées (en cas de reconstruire l'arbre) et choisir l'entrée dont la distance est la plus petite.
Absorb	Absorber 2 entrées. Ces deux entrées sont l'entrée à insérer dans l'arbre et son entrée la plus proche dans le nœud. L'entrée à insérer est absorbée par l'entrée la plus proche si la compacité de la sous-classe formée en fusionnant ces deux entrées est inférieure au seuil d'absorption.
insertEnd	Insérer une entrée à la fin de la liste d'entrée du nœud (en cas de ne pas absorber une entrée)
testSplit	Prévoir si une insertion à la fin d'une entrée au nœud peut causer le dépassement de capacité de l'arbre. C'est-à-dire calculer le nombre de nœuds qui sont créés par une insertion d'entrée pour décider quand il faut reconstruire l'arbre.
Split	Décomposer le nœud en deux nœuds en cherchant tout d'abord le pair l'entrée qui sont les plus loin et puis, distribuant toutes les entrées restantes au tour de ces entrées selon le critère de proximité.
updateParent	Mettre à jour le contenu de l'entrée parent si le nœud insère, absorbe ou supprime une entrée.
insertOutlier	Insérer une entrée qui est considérée comme un aberrant potentiel sauvegardé dans le disque.
computePath	Calculer le chemin du nœud qui sert à reconstruire l'arbre. Un chemin est une collection des étiquettes des entrées à partir du nœud de racine au nœud actuel. L'étiquette d'une entrée est sa position dans un nœud.
Reinsert	Réinsérer une entrée de feuille lors de la reconstruction de l'arbre.

Pour insérer une entrée E à un nœud, on cherche l'entrée la plus proche de E. Si le nœud actuel est une feuille, on vérifie si E peut être absorbée par son entrée la plus proche en fusionnant ces deux entrées pour former une sous-classe et teste si la compacité (le diamètre) de la sous-classe formée est inférieure au seuil d'absorption.

S'il ne l'est pas, on calcule le nombre de nœuds qui seront créés si on insère E à la fin de liste d'entrées du nœud. Si le nombre de nœuds créés dépasse le nombre de nœuds restant, alors on stocke E sur le disque comme un point retardé. S'il y a assez d'espace, alors on insère E au nœud. Si le nœud actuel n'est pas une feuille, alors on descend au nœud fils de l'entrée la plus proche de E.

Si E est insérée à la fin de la liste d'entrées du nœud et le nœud dépasse sa capacité, alors il faut scinder ce nœud en 2 nœuds en cherchant deux entrées qui sont les plus loin et distribuer les entrées restantes au tour ces deux germes selon le critère de proximité : On affecte une entrée à un germe si elle est plus de ce germe par rapport au germe restant. On crée une nouvelle entrée qui joue le rôle de parent de nouveau nœud et insère cette nouvelle entrée au nœud parent (qui contient l'entrée parent du nœud divisé). La division peut être propagée en haut de l'arbre. Si la racine est divisée, alors la hauteur de l'arbre est augmentée par 1.

L'insertion, la division et l'absorption évoquent la mise à jour des entrées parent de bas en haut.

Classe CCFEntryList

Pour construire une liste d'entrée, il faut spécifier sa taille. C'est également la taille du nœud qui possède cette liste.

Les opérations principales d'une liste d'entrées :

Nom	Description
insert	Insérer une entrée à la fin de la liste d'entrées.
remove	Supprimer une entrée dans la liste à n'importe quelle position.

Classe CCFEntry

En étant donné un seuil d'absorption, on crée une entrée contenant un vecteur de CF nul. Le fils d'une entrée est le nœud auquel il pointe, autrement dit, c'est la sous-classe que cette entrée contient. Chaque entrée a une enveloppe qui est le nœud la contient une fois cette entrée est insérée à ce nœud.

Classe CCFVector

C'est une classe vide qui supporte les opérations de calculer la distance, la compacité, l'addition et la soustraction d'une entrée décrite ci-dessus.

CCFNumVector

C'est une classe qui hérite la classe CCFVector. Elle se compose du nombre des points, de la somme linéaire des points et de la somme des carrées des points dans une sous classe. Nous avons utilisé cette classe pour la classification des données purement numériques.

Publications de l'auteur

- Sur le sujet du doctorat :

- [1] A. Chaari, S. Lelandais, M. B. Ahmed, "*Face classification scheme simplifying identification in large scale biometric databases*", Fourth IEEE International Conference on Systems, Signals & Devices (SSD), Hammamet, Tunisie, 19-22 mars 2007.
- [2] M. Saaidia, A. Chaari, S. Lelandais, V. Vigneron, M. Bedda, "*Face localization by neural networks trained with Zernike moments and Eigenfaces feature vectors. A comparison*", IEEE Int. Conf. on Advanced Video and Signal based Surveillance (AVSS), London, September 2007.
- [3] A. Chaari, M. B. Ahmed, S. Lelandais, "*Appearance based face identification and classification. A Combining approach*", The 3rd International Conference on Information & Communication Technologies: from Theory to Applications (ICTTA), Damascus (Syria), April 2008.
- [4] D. P.-Delacrétaz, S. Lelandais, J. Colineau, L. Chen, B. Dorizzi, E. Krichen, M. A.-Mellakh, A. Chaari, S. Guerfi, M. Ardabilian, J. D'Hose, B. Benamor, "*The IV2 Multimodal Biometric Database (Including Iris, 2D, 3D, Stereoscopic, and Talking Face Data), and the IV2-2007 Evaluation Campaign*", IEEE Second International Conference on Biometrics Theory, Applications and Systems (BTAS), USA, 2008.
- [5] A. Chaari, S. Lelandais, M. Saaidia, V. Vigneron, M. B. Ahmed, "*Comparaison d'approches pour la détection et l'identification de visages 2D*", 14ème Colloque National de la Recherche en IUT (CNRIUT), Lyon, France, Juin 2008.
- [6] A. Chaari, S. Lelandais, M. B. Ahmed, "*A New Clustering Approach for Face Identification*", Image Processing Theory, Tools and Applications (IPTTA), Sousse, Tunisie, pp. 1-6, Novembre 2008.
- [7] A. Chaari, S. Lelandais, M. B. Ahmed, "*A Pruning Approach Improving Face Identification Systems*", accepté à IEEE Int. Conf. on Advanced Video and Signal based Surveillance (AVSS), Genoa, Italie, September 2-4, 2009.
- [8] A. Mellakh, A. Chaari, S. Guerfi, J. Dhose, J. Colineau, Sylvie Lelandais, D. P.-Delacrétaz, B. Dorizzi, "*2D Face Recognition in the IV2 Evaluation Campaign*", Advanced Concepts for Intelligent Vision Systems (ACIVS), Bordeaux, France, 2009.
- [9] A. Chaari, S. Lelandais, M. B. Ahmed, "*Face classification scheme simplifying identification in biometric databases*", Transactions on Systems, Signals & Devices (TSSD), Shaker-Verlag, sous presse, 2009.

- Autres publications:

- [10] C. Montagne, S. Lelandais, A. Chaari et M. B. Ahmed, "*Invariant couleur basé sur la transformation du boulanger – Application à la localisation globale d'un robot d'intérieur*", Int. Conf. Sciences of Electronics, Technology of Information and Telecommunication (SETIT), Tunisie, Mars 2005.
- [11] S. Lelandais, A. Chaari, A. Smolarz, C. Montagne, B. Jacquin, "*A new color invariant for image retrieval using the baker's transformation*", Beijing International Conference on Imaging (BICI), China, May 2005.
- [12] A. Chaari, S. Lelandais, C. Montagne et M. Ben Ahmed, "*Color image retrieval techniques for a global localization of an indoor mobile robot*", 3rd European Conference on Color in Graphics, Imaging and Vision (CGIV), Leeds, UK, June 2006.
- [13] A. Chaari, S. Lelandais, C. Montagne et M. Ben Ahmed, "*Global interior robot localization by a color content image retrieval system*", EURASIP Journal on Advances in Signal Processing, accepté en Aout 2007, 15 pages, vol. 2008.

