



Thèse de doctorat préparée au Département Électronique et Physique de
l'Institut National des Télécommunications dans le cadre de l'École Doctorale
SITEVERY en co-accréditation avec l'université d'Evry-Val d'Essonne

Thèse N°
00000000000

THÈSE

pour l'obtention du grade de Docteur
de **L'INSTITUT NATIONAL DES TÉLÉCOMMUNICATIONS**

Spécialité : **Traitement d'image et de signal**

Présentée et soutenue publiquement par:

Anouar Mellakh

Titre:

Reconnaissance des visages en conditions dégradées

Soutenue le 07 Avril 2009 devant le jury composé de :

Président du jury	M. Maurice Milgram
Rapporteur	M. Faouzi Gorbel
Rapporteur	Mme. Sylvie Lelandais
Examineur	M. André Gagalowicz
Directeur de thèse	Mme. Bernadette Dorizzi
Encadrant de thèse	Mme. Dijana Petrovska-Delacrétaz

À ma femme Tiphaine
À mes parents
À mon frère et ma soeur
En souvenir de Christine

Résumé

De nos jours, les algorithmes de reconnaissance du visage proposés dans la littérature ont atteint un niveau de performance correct lorsque les conditions d'acquisition des images comparées sont contrôlées, mais les performances chutent lorsque ces conditions sont dégradées. Les conditions d'acquisition contrôlées correspondent à un bon équilibre d'illumination, ainsi qu'une haute résolution et une netteté maximale de l'image du visage.

Afin de bien cerner le problème de dégradation des performances dans les conditions d'acquisition difficiles et de proposer par la suite des solutions adaptées, nous avons effectué plusieurs études à différents niveaux de la chaîne de la reconnaissance. Ces études concernent le comportement des algorithmes basés sur les approches globales. Elles concernent également les différentes méthodes de normalisation photométrique ainsi que des stratégies de reconnaissance basées sur la qualité des images du visage.

Les solutions proposées à chaque niveau de cette chaîne ont apporté une amélioration significative des performances par rapport aux approches classiques. Pour les algorithmes de reconnaissance, nous avons proposé l'utilisation de la fusion de la phase et de l'amplitude des représentations de Gabor de l'image du visage comme nouveau gabarit, à la place de l'image des pixels. Bien que les représentations de Gabor aient été largement utilisées, particulièrement dans les algorithmes basés sur les approches globales, la phase n'a jamais été exploitée jusqu'à ce jour. Nous expliquons dans cette thèse les problèmes liés à l'utilisation de cette phase de Gabor et nous proposons une solution pour y remédier.

Différentes méthodes de normalisation photométrique de l'image du visage ont été étudiées et comparées. Nous avons, par la suite, proposé une nouvelle approche de normalisation basée sur la correction de la composante luminance. Enfin, nous

avons présenté une stratégie de reconnaissance se basant sur la mesure de qualité du visage. Cette mesure est une fusion de plusieurs critères de qualité et selon nos expériences, cette stratégie permet une amélioration notable des performances par rapport aux méthodes classiques.

Les différentes études, la validation de nos mesures de qualité ainsi que la validation de la stratégie de reconnaissance ont été effectuées sur les deux bases de données publiques de visage FRGCv2 et BANCA largement utilisées par la communauté.

Mot-clefs

Biométrie, Reconnaissance de visage, Illumination, Normalisation photométrique, filtres de Gabor, Qualité.

Summary

Nowadays, the algorithms of face recognition, proposed in the literature, reached a correct performance level when the acquisition's conditions for the tested images are controlled, but this performances fall when these conditions degraded. The controlled conditions of acquisition correspond to a good balance of illumination, as well as a high-resolution and a maximum sharpness of the face image.

In order to determine the problem of degradation of performances under difficult capture's conditions and also to propose adapted solutions, we carried out several studies at various levels of the recognition's chain. These studies relate to the behavior of the algorithms based on global approaches. They also relate to the various methods of photometric standardization as well as strategies of recognition based on the quality of the face images.

The solutions suggested on each level of this chain resulted in a significant improvement of the performances compared to the traditional approaches. For the recognition algorithms, we proposed to fuse the phase and magnitude of Gabor's representations of the face as a new representation, in the place of the raster image. Although the Gabor representations were largely used, particularly in the algorithms based on global approaches, the Gabor phase was never exploited. We explain in this thesis the problems involved in the use of this phase and we propose a solution to solve this problem. Various methods of photometric normalization for face were studied and compared. We, thereafter, proposed a new approach of normalization based on the correction of the brightness component. Lastly, we presented a strategy of recognition based on the quality measure of face. This measurement is a fusion of several quality standards and according to our experiments; this strategy offers an improvement of the verification rate

compared to the classical methods.

The various studies, the validation of our quality measurements as well as the validation of the recognition strategy were carried out on the two public and largely used databases of FRGCv2 face and BANCA.

Keywords

Biometrics, Face Recognition, Illumination, Photometric Normalisation, Gabor filters, Quality.

Table des matières

Table des matières	1
Table des figures	5
Liste des tableaux	9
1 Introduction	13
1.1 Biométrie du visage	14
1.2 Variabilité de la modalité visage	16
1.3 Motivation et contributions	18
1.4 Structure de la thèse	19
2 État de l'art de la reconnaissance de visage	21
2.1 Introduction	21
2.2 État de l'art des techniques de reconnaissance	23
2.2.1 Méthodes de réduction d'espace	24
2.2.2 Approches par points caractéristiques	27
2.3 Influence des variations environnementales sur la reconnaissance par le visage	29
2.4 Les approches utilisant le filtrage de Gabor	34
2.5 Conclusion	35
3 Outils d'évaluation, bases de données et protocoles	37
3.1 Introduction	37
3.2 Scores et protocoles d'évaluation des systèmes biométriques	40
3.3 Évaluation des systèmes de reconnaissance	42
3.3.1 Évaluation des systèmes de vérification biométrique	42

3.3.2	Évaluation des systèmes d'identification biométrique	49
3.4	Bases de données publiques	51
3.4.1	Face Recognition Grand Challenge [NIST]	51
3.4.2	BANCA [SURREY]	55
3.4.3	IV ²	57
3.5	Conclusion	60
4	Les algorithmes de reconnaissance de visage	61
4.1	Introduction	61
4.2	Dimensionnalité de l'espace visage	62
4.3	Analyse en composantes principales	63
4.3.1	Réduction d'espace par ACP	64
4.3.2	Application de l'ACP à la reconnaissance de visage	67
4.4	Fisher Discriminant Analysis	72
4.4.1	Analyse Linéaire Discriminante	73
4.4.2	La LDA directe (DLDA)	74
4.5	Tests et comparaisons des algorithmes	76
4.6	Conclusion	86
5	Normalisation de l'illumination	87
5.1	Introduction	87
5.2	Normalisation géométrique	88
5.3	Extraction de l'intensité du pixel de l'espace couleur	90
5.4	Normalisation de l'illumination	93
5.4.1	Égalisation d'histogramme	93
5.4.2	Correction Gamma	94
5.4.3	MultiScale-Retinex	96
5.4.4	Lissage anisotropique	97
5.4.5	Nouvelle méthode de correction de l'image par transforma- tion de l'illumination	99
5.5	Résultats et analyses	105
5.5.1	Comparaison entre les différentes transformations en ni- veaux de gris	105
5.5.2	Comparaison des différentes méthodes de normalisation . .	106
5.6	Conclusion	107

6	Analyse globale par filtrage de Gabor	109
6.1	Introduction	109
6.2	Filtrage de Gabor et extraction des caractéristiques	110
6.3	Représentation du visage et choix des filtres de Gabor	112
6.3.1	Influence des caractéristiques de la famille des filtres de Gabor sur les performances de la reconnaissance	113
6.3.2	Études des performances suivant le choix des représentations de Gabor	116
6.3.3	Fusion de l'amplitude et de la phase des représentations de Gabor	119
6.3.4	Validation de la fusion de l'amplitude et de la phase corrigée sur des algorithmes de réduction d'espace à base de noyau	120
6.4	Résultats sur la base FRGC	121
6.5	Résultats sur d'autres bases	122
6.5.1	Résultats sur la base BANCA	123
6.5.2	Résultats sur la base IV ²	123
6.6	Conclusion	124
7	Mesures de qualité pour le visage	127
7.1	Mesures de qualité	130
7.1.1	Mesure du déséquilibre de l'illumination par distributions locales	130
7.1.2	Mesure de la qualité de l'illumination par SVM	133
7.1.3	Mesure de la netteté	136
7.1.4	Classification par fusion des mesures de qualité	140
7.1.5	Validation des mesures de qualité sur la base de développement	143
7.2	Qualité et stratégie de reconnaissance	144
7.2.1	Protocoles et bases de données	145
7.2.2	Scénario : références Passeport	147
7.2.3	Scénario : références dégradées	153
7.2.4	Récapitulatif des tests	157
7.3	Validation de la stratégie sur la base BANCA	159
7.4	Conclusion	159
8	Conclusion et Perspectives	161

Bibliographie	169
----------------------	------------

Table des figures

1.1	Exemples de variation d'illumination de la base YaleB [24]	16
1.2	Exemple de réflexion diffuse de la lumière (a) et de réflexion spéculaire (b)	17
2.1	Exemples des 5 premiers visages propres construits avec l'approche ACP (a) et Exemples des 5 premiers visages propres construits avec l'approche LDA (b)	26
2.2	Structure des approches EGM et EBGM	28
3.1	Exemple de distributions théoriques des scores "intra-classes" et "inter- classes" pour un système biométrique théoriquement parfait	41
3.2	Exemple de distributions théoriques des scores "intra-classes" et "inter- classes" pour un système biométrique réel	42
3.3	Variation des taux FAR, FRR et TER en fonction du seuil de décision	44
3.4	Exemple de courbes ROC (Source NIST reconnaissance de la parole)	47
3.5	Exemple de courbes DET (Source NIST reconnaissance de la parole)	47
3.6	Courbe de rang en fonction du taux de bonne identification (Source NIST évaluation du visage FERET)	51
3.7	Exemple de sessions d'acquisition de la base de données FRGCv2	52
3.8	Premiers résultats obtenus lors des expériences de visage 2D de la base FRGCv2 (Source NIST Biometrics Consortium Conference)	54
3.9	Exemples d'image de la base de données BANCA selon trois scénarios : (a) contrôlé, (b) dégradé et (c) non contrôlé	55
3.10	Exemples de données de la base IV ²	59

3.11	Exemples d'acquisition d'images 2D : (a) image haute résolution et (b) image dégradée	60
4.1	Exemple de distribution d'un nuage de points dans un espace 2D	64
4.2	Exemple d'image de visage normalisé (nombre de pixels fixe)	66
4.3	Évolution de la mesure du taux d'énergie résiduelle en fonction de la variance de l'espace pour la sous-base apprentissage de FRGC	69
4.4	Évolution du EER en fonction de la variance pour l'expérience 1 et l'expérience 4 de FRGC	70
4.5	Évolution du EER en fonction de la variance pour le protocole P de BANCA	70
4.6	Comparaison entre les projections de deux classes de points ("class 1" et "class 2") sur les axes principaux construits par ACP et par LDA	74
4.7	Diagramme de l'algorithme de la LDA directe (DLDA)	76
4.8	Exemple d'image de brillance normalisée géométriquement	78
4.9	Courbes ROC des algorithmes ACP, LDA et DLDA pour les expériences 1 et 4 de FRGC : avec égalisation d'histogramme, base d'apprentissage BA	81
4.10	Courbes ROC des algorithmes ACP, LDA et DLDA pour les expériences Mc et P de la base BANCA : avec égalisation d'histogramme, base d'apprentissage BA	82
4.11	Comparaison des performances pour les algorithmes ACP, LDA et DLDA pour l'expérience 1 sur la base FRGC	83
4.12	Comparaison des performances pour les algorithmes ACP, LDA et DLDA pour l'expérience 4 sur la base FRGC	84
4.13	Variation des performances de la DLDA en fonction de la variation du nombre d'images acquises en conditions non contrôlées	85
4.14	Variation des performances de la DLDA en fonction de la variation proportionnelle du nombre d'images acquises en conditions contrôlées et non contrôlées	85
5.1	Exemple d'une normalisation géométrique de visage de la base FRGC	88
5.2	Variation du taux de EER de l'algorithme de la DLDA en fonction de la normalisation géométrique sur la base FRGC pour l'expérience 1	89
5.3	Différents systèmes colorimétriques	91

5.4	Différentes transformations de l'image couleur d'un exemple de la base FRGC	92
5.5	Exemple d'égalisation d'histogramme	94
5.6	Schématisation de la réflexion diffuse	96
5.7	Discrétisation au voisinage d'un pixel	98
5.8	Normalisations photométriques du visage	99
5.9	Schéma des modèles de l'illumination : (a) modèle de l'illumination originale, (b) modèle d'illumination après alignement	101
5.10	Décomposition de l'illumination référence en somme de deux illuminations verticalement symétriques.	101
5.11	Exemple d'une illumination originale en (a) et d'une illumination symétrique en (b)	102
5.12	Exemples de correction de l'illumination par symétrie axiale	103
5.13	Correction de l'image du visage par transformation de l'illumination .	104
5.14	Courbes ROC de l'expérience 1 (a) et de l'expérience 4 (b) de la base FRGC	107
6.1	Partie réelle (a) et imaginaire (b) du filtre de Gabor	111
6.2	Parties réelles (a) et imaginaires (b) du filtre de Gabor à 4 niveaux de résolution et selon 4 orientations	111
6.3	Résultats de la convolution d'une image de visage avec une famille de 16 filtres de Gabor (4 orientations (horizontales) et 4 résolutions (verticales))	112
6.4	Mesure de similarités entre les axes principaux de la DLDA construits par les 4,5 et 6 premiers niveaux	116
6.5	Évolution des taux de EER (a) et de VR@0.1% de FAR (b) en fonction du seuil Th pour le protocole PEG	118
6.6	Courbes ROC des différentes représentations de Gabor pour le protocole PEG avec une famille de filtres à 4 résolutions et 8 orientations .	119
6.7	Courbes ROC des différentes méthodes pour le protocole PEG	121
7.1	Principe de calcul d'une mesure de qualité globale ($Q1_a$ ou $Q1_b$) à partir des mesures locales (moyennes ou médianes, respectivement) .	131
7.2	Distributions de l'illumination par mesure des moyennes locales, $Q1_a$, pour les bases [BQC] et [BQN]	132

7.3	Distributions de l'illumination par mesure des médianes locales, $Q1_b$, pour les bases [BQC] et [BQN]	133
7.4	Exemples des histogrammes des images de type contrôlé (A) et de type non contrôlé (B)	134
7.5	Variations du taux de classification correcte pour le classifieur SVM à noyau polynomial suivant le rang du polynôme sur la base de développement	135
7.6	Variations du taux de classification correcte pour le classifieur SVM à noyau radial sur la base de développement	135
7.7	Distributions des mesures de netteté $Q3_a$ de [BQC] et de [BQN] par méthode du gradient	137
7.8	Mesure du flou par approche fréquentielle locale pondérée	138
7.9	Distributions des mesures de netteté $Q3_b$ de [BQC] et de [BQN] par l'approche fréquentielle globale	139
7.10	Distributions des mesures de netteté $Q3_c$ de [BQC] et de [BQN] par l'approche fréquentielle locale pondérée	139
7.11	Évolution de l'inertie intra-classes en fonction du nombre des centroïdes pour la base d'apprentissage	142
7.12	Distribution de la mesure q pour les images de la base de développement [[BQC] + [BQN]]	143
7.13	Exemples d'images de type contrôlé (A), de type dégradé (B) et de type non contrôlé (C) de la base de données BANCA	143
7.14	Découpage de la base FRGCv2	147
7.15	Courbes ROC de Test1	148
7.16	Courbes ROC de Test2	149
7.17	Courbes ROC de Test3	150
7.18	Courbes ROC de Test4	150
7.19	Courbes ROC de Test5	152
7.20	Courbes ROC de Test6	152
7.21	Courbes ROC de Test7	154
7.22	Distribution des scores intra-classes (A) et Distribution des scores inter-classes (B)	155
7.23	Courbes ROC des Tests 8-9 et 10 (D1, D2 et D3)	157

Liste des tableaux

2.1	Liste des participants aux campagnes FRVT	22
3.1	Nombre de tests pour les 3 protocoles dans les expériences 1, 2 et 4 de la base FRGC	53
3.2	Résultats du protocole P sur la base BANCA	57
3.3	Protocoles d'évaluation pour la base IV ²	58
4.1	Caractéristiques des sous-ensembles d'apprentissage de l'espace de réduction par ACP	72
4.2	Résultats en EER de l'expérience 4 de FRGC pour les différentes bases d'apprentissage	72
4.3	Résumé des résultats aux points de fonctionnement (EER et VR@0.1% de FAR) des 3 algorithmes (ACP, LDA, DLDA) pour les deux types d'expérience sur les bases FRGC et BANCA	80
5.1	Résultats de la DLDA donnés pour les différentes méthodes d'extraction des niveaux de gris (RVB, TSL, TSV) sur la base FRGC aux deux points de fonctionnement EER et VR@0.1% de FAR (Expérience 1)	105
5.2	Résultats de la DLDA donnés pour les différentes méthodes d'extraction des niveaux de gris sur la base BANCA au point de fonctionnement EER (Protocole Mc)	105
5.3	Résultats de reconnaissance de la DLDA appliquée à différentes méthodes de pré-traitement. Les résultats sont présentés au point de fonctionnement VR @0.1% de FAR sur la base FRGC pour les expériences 1 et 4.	106

5.4	Résultats de reconnaissance de la DLDA appliqué à différentes méthodes de pré-traitement. Les résultats sont présentés au point de fonctionnement EER sur la base BANCA pour les protocoles Mc et P.	106
6.1	Performances en EER [\pm CC] pour chaque niveau de résolution en fonction du nombre d'orientations choisi par le filtrage de Gabor sur le protocole PEG	114
6.2	Résultats en EER et en VR@0.1% de FAR des combinaisons cumulatives des différents niveaux de résolution avec 8 orientations pour le protocole de test PEG	115
6.3	Moyenne et écart-types des similarités entre les axes principaux de la DLDA construits par les 4, 5 et 6 premiers niveaux	115
6.4	Résultats en EER et en VR@0.1% de FAR des différentes représentations de Gabor	117
6.5	Résultats en EER et en VR@0.1% de FAR des différentes fusions des représentations de Gabor	120
6.6	Résultats de la reconnaissance par fusion amplitude-phase en taux de VR@0.1% de FAR pour les deux expériences 1 et 4 de la base FRGC ainsi que quelques résultats de la littérature	122
6.7	Résultat en WER (<i>Weighted Error Rate</i>) de la DLDA appliquée à la fusion amplitude-phase de réponses de la famille de 32 filtres de Gabor sur la base BANCA (protocole P)	123
6.8	Protocoles d'évaluation pour la base IV ²	123
6.9	Taux de EER pour les différents algorithmes sur la base IV ² (visage 2D) participant à la première évaluation	124
7.1	Taux de Bonne Classification (TBC) pour chaque type de noyau pour les deux bases [BQC] et [BQN]	136
7.2	Taux d'Erreur de Classification (TEC) suivant la méthode de mesure du flou	140
7.3	Taux de Classification Correcte (TCC) pour les deux classes sélectionnées de la base BANCA avec notre mesure q	144
7.4	Tableau récapitulatif des divers scénarios et protocoles avec les résultats des performances donnés en VR@0.1% de FAR	158

7.5	Résultats de la stratégie de reconnaissance sur la base BANCA, Protocole P et G, utilisant l'algorithme de reconnaissance DLDA appliqué aux représentations Gabor du visage.	159
-----	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----

Chapitre 1

Introduction

Dans la civilisation babylonienne, au VI^{ème} siècle avant J.-C., l’empreinte du pouce laissée sur une poterie d’argile permettait de sceller des accords commerciaux. Les égyptiens utilisaient les descriptions physiques des commerciaux pour différencier ceux qui sont connus de ceux qui sont nouveaux sur le marché. Ces exemples montrent que les caractéristiques physiques ont toujours été utilisées par l’Homme, bien avant l’avènement des sciences modernes et des ordinateurs, pour identifier les personnes.

En réalité la Biométrie, comme elle est connue de nos jours, découle directement de l’anthropométrie moderne. Cette technique permet de reconnaître des personnes à partir de leurs caractéristiques physiques et comportementales. L’anthropométrie a vu le jour au XIX^{ème} siècle avec Alphonse Bertillon (1853-1914). Ce criminologue français fonda en 1870 le premier laboratoire de police scientifique d’identification criminelle et inventa l’anthropométrie judiciaire appelée "système Bertillon" ou "bertillonnage". Le système de reconnaissance des personnes, proposé par Bertillon, est basé sur la mesure de différents paramètres du corps humain [8]. Dans ses travaux, Bertillon affirme qu’en prenant quatorze mensurations (taille, pieds, mains, nez, oreilles, etc.) sur n’importe quel individu, il n’y a qu’une seule chance sur deux cent quatre-vingt-six millions pour qu’on retrouve les mêmes mesures chez une autre personne. Ce système a été adopté rapidement dans toute l’Europe, puis aux États-Unis.

La première utilisation des empreintes digitales comme moyen d’identification remonte à 1877 aux Indes où le britannique William Herschel eut l’idée de les

utiliser pour éviter que les bénéficiaires de pension de l'armée ne la touchent plusieurs fois. À cette époque, les empreintes digitales servent aussi à authentifier des documents officiels. Sir Francis Galton améliora cette technique et publia des travaux détaillés sur la reconnaissance par l'empreinte digitale, basée sur des caractéristiques particulières de la texture, appelées les minuties [13]. Suite à ces publications, Scotland Yard a ouvert le premier fichier d'empreintes digitales en 1901, sous la direction du commissaire Edward Henry. Ce procédé émergea ensuite en Asie, en Afrique du Sud et en Europe. La première utilisation des empreintes digitales en France fut le 24 octobre 1902 pour confondre le criminel Henri-Léon Scheffer, deux ans après sa disparition. Aux Etats-Unis, les prisons commencèrent à rendre obligatoire l'enregistrement des empreintes des détenus dès 1903.

Alors que la méthode de Bertillon fût abandonnée à partir de 1970, l'enregistrement systématique des empreintes connut en revanche un essor mondial.

Dès le début des années 60, différents moyens physiques et comportementaux, permettant une reconnaissance de l'individu, ont été utilisés. Parmi ces moyens, communément appelés "modalité biométrique", on trouve le visage, l'empreinte digitale, l'iris et la forme de la main. Récemment, d'autres modalités ont été introduites comme les veines de la main, le réseau veineux de la rétine, l'empreinte de l'oreille et l'ADN. Pour ce qui est des modalités comportementales, on peut citer la signature (dynamique ou statique), la démarche...

Avec l'avènement des ordinateurs, l'idée d'une reconnaissance automatique est née ; c'est le début de la biométrie moderne. Depuis les années 60, plusieurs travaux sur des algorithmes automatiques ou semi-automatiques furent publiés ; citons le cas du visage, de la signature, de la voix et bien sûr, des empreintes digitales. Grâce à cet élan de recherche et ces travaux très prometteurs, le FBI lança en 1969 une campagne scientifique afin de développer un système automatique de reconnaissance par les empreintes digitales. La géométrie de la main, dont les balbutiements remontent au milieu du XIXème siècle, connut dans les années 70 un regain d'intérêt avec la commercialisation du premier produit biométrique de contrôle d'accès et d'identification.

1.1 Biométrie du visage

De nos jours, le visage peut être utilisé pour identifier une personne dans une base mais il est plus communément utilisé pour vérifier l'identité. Il s'agit alors

de déterminer si une identité réclamée est correcte ou fausse. Pour la vérification des visages, ce processus est effectué en comparant un modèle du demandeur (une ou plusieurs images de test), avec un modèle stocké (une ou plusieurs images de référence). Le processus complet de vérification des visages est décrit ci-dessous :

- Capture du visage : le visage d’une personne est numérisé par des capteurs comme des appareils photo numériques pour la représentation 2D du visage, des scanners 3D pour avoir l’information de profondeur de l’image combinée avec l’image de la texture ou aussi des capteurs infrarouge ou proche infrarouge. Des informations spécifiques à la capture peuvent être déterminées aussi lors de l’acquisition de l’image, comme la pose de la personne par rapport à la caméra, l’expression ou aussi les conditions d’illumination de la capture de l’image. Dans cette thèse nous nous consacrons à l’étude exclusive du problème de la reconnaissance des visages 2D.
- Détection : L’image 2D acquise peut contenir à la fois le visage de la personne et éventuellement, un arrière-plan. Dans le processus de détection, le visage est localisé dans l’image. Souvent, la sortie de cette étape est la localisation des centres des yeux.
- Normalisation : La normalisation est constituée de deux processus : géométrique et photométrique. La normalisation géométrique est nécessaire parce que la taille du visage à l’intérieur de l’image acquise peut varier en fonction de la distance entre le module d’acquisition et la personne. Le visage doit donc être extrait à partir de l’image et une transformation géométrique, pour obtenir une taille fixe, est appliquée. L’approche standard consiste à définir l’emplacement des centres des yeux dans une position constante au sein de l’image de sortie.
L’étape de normalisation photométrique tente d’éliminer ou de réduire les effets de l’illumination de l’image. Dans certains cas, l’étape de normalisation photométrique peut être appliquée avant, ou avant et après l’étape de normalisation géométrique. Elle peut aussi être appliquée pendant la phase de détection.
- Calcul de score : L’image normalisée est ensuite comparée à une ou plusieurs images stockées (si un modèle utilisant une galerie d’images est utilisé). Cette comparaison produit un score, représentant la similarité ou la distance de l’image test par rapport à l’image ou au modèle référence.
- Décision : Le score est comparé à un seuil qui détermine si l’image est

acceptée ou rejetée en tant que représentant de l'identité proclamée.

1.2 Variabilité de la modalité visage

La différence d'apparence d'un même visage capturé dans deux conditions d'acquisition distincte pose un énorme problème dans le domaine de la reconnaissance faciale. Cette différence est dû, généralement, à des facteurs d'environnement comme les conditions d'éclairage, les caractéristiques des capteurs et aussi leur positionnement par rapport au visage lors de l'acquisition. Cette variation peut aussi être dûs aux modifications du visage liées aux expressions ou aux changements de poids ainsi qu'à l'âge.

Généralement, il est admis que la distance entre deux visages de personnes différentes acquis dans des conditions propres est plus faible que celle qui existe pour une même personne dans des conditions variables d'éclairage et de mise au point .



FIGURE 1.1: Exemples de variation d'illumination de la base YaleB [24]

Comme le montre la figure 1.1, une variation de l'éclairage peut sérieusement altérer l'apparence d'un visage dans l'image, dans la mesure où les images à l'extrême droite semblent plus proches les unes des autres que de leurs correspondants avec un éclairage frontal (extrême gauche).

Une formalisation des observations qualitative précédentes, on peut remarquer qu'il existe deux formes de réflexion de la lumière par une surface. La réflexion peut être spéculaire ou bien diffuse, suivant la nature de la surface (Figure 1.2). Les lois géométriques de la réflexion ne s'appliquent qu'à la réflexion spéculaire ;

il faut faire appel à des modélisations plus complexes pour traiter la réflexion diffuse.

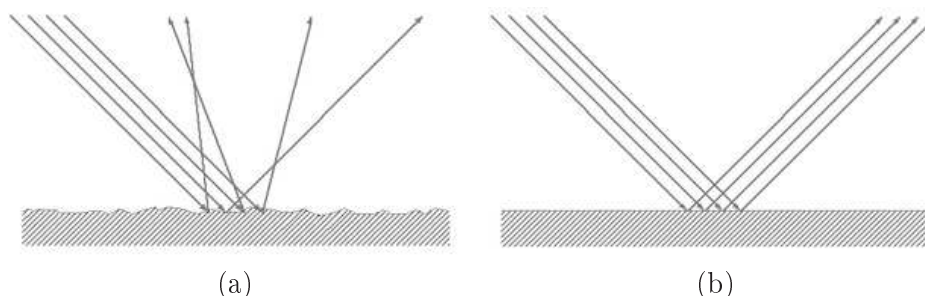


FIGURE 1.2: Exemple de réflexion diffuse de la lumière (a) et de réflexion spéculaire (b)

La réflexion diffuse intervient sur les interfaces irrégulières, la lumière est réfléchie dans un grand nombre de directions et l'énergie du rayon incident est redistribuée dans une multitude de rayons réfléchis.

La réflexion est dite spéculaire lorsque le rayon incident donne naissance à un rayon réfléchi unique. Idéalement, l'énergie du rayon incident se retrouve totalement dans le rayon réfléchi, en pratique une partie de l'énergie peut être absorbée ou diffusée au niveau de l'interface.

Ces deux types de réflexion peuvent donner lieu à un effet indésirable de l'illumination : les ombres. On distingue deux types d'ombres, les ombres propres et les ombres portées :

- Une ombre propre se produit lorsque la lumière arrive sur une surface avec un changement de pente. La luminosité des pixels correspondante à cette zone diminue à mesure que l'angle d'incidence s'écarte de la normale de la surface. Cette luminosité atteint son minimum lorsque l'angle entre la lumière incidente et la normale de la surface est égale à 90°).
- Une ombre portée se produit lorsque la source de lumière est occultée par un objet avant la réflexion de la lumière sur la surface.

En traitement d'images, une ombre est considérée comme étant une région à faible luminosité et ayant des contours à fort gradient.

Dans cette thèse j'étudie et je développe particulièrement l'aspect variation d'illumination et je propose des solutions à divers niveaux de la chaîne de traitement qui permettront d'améliorer les performances de reconnaissance.

1.3 Motivation et contributions

Les systèmes de reconnaissance par le visage ont atteint pendant cette dernière décennie des performances élevées lorsque les images de référence et de test sont acquises dans des conditions propres. Ces performances se dégradent de façon significative lorsque les conditions d'acquisition sont très variables. Ceci a été constaté dans plusieurs campagnes d'évaluation internationales (FRGC 2005 , FRVT 2006 et MBGC 2008).

En réalité, l'information (qui peut être considérée comme du bruit), introduite par l'effet de la variation de l'illumination ou des conditions d'acquisition, ne peut pas être facilement séparée de l'information discriminante entre les visages. Diverses solutions ont été proposées dans la littérature afin d'atténuer l'impact de ce bruit sur les performances de la reconnaissance.

C'est dans ce contexte de variabilité d'illumination qu'intervient cette thèse. Nous avons proposé de nouveaux algorithmiques alliés à des stratégies originales pour améliorer les résultats de la reconnaissance par le visage, dans ce contexte. En effet, même si plusieurs solutions pour réduire l'influence des effets de l'illumination ont déjà été proposées dans la littérature elle se insuffisantes dans la pratique.

Nous nous sommes intéressés tout d'abord à des méthodes permettant de corriger les effets d'illumination et nous avons proposé un algorithme original dans ce cadre. Nous avons aussi proposé un algorithme de reconnaissance de visage original que nous avons comparé à différents algorithmes de la littérature sur des images présentant des illuminations variables issues des grandes bases de données largement diffusées dans la communauté biométrique.

Des études ont montré aussi que des critères de qualité comme la netteté de l'image, le contraste, l'illumination ou bien la pose et l'expression sont des facteurs qui influencent, à divers degrés, les performances des systèmes de reconnaissance [12,22]. Ainsi, lors du "*Biometric Quality Workshop*" en 2006, Werner et Brauckmann ont montré la corrélation entre les différents critères de qualité (netteté, contraste et pose) et les résultats des systèmes de reconnaissance. L'étude a été effectuée sur la base FRGCv2 et sur deux types d'algorithmes, un algorithme basé sur une approche globale et un deuxième sur une approche par points caractéristiques [12]. Lors du même workshop, Weber [98] a également présenté les résultats d'une étude semblable à celle de Werner et il est arrivé aux mêmes conclusions de corrélation.

En nous basant sur ces constatations, nous avons introduit de nouvelles mesures de qualité sur les visages, principalement orientées vers la mesure de l'illumination et du flou et nous avons proposé par la suite de nouvelles stratégies de reconnaissance basées sur ces mesures de qualité.

1.4 Structure de la thèse

Cette thèse se compose de 6 grands chapitres et elle est organisée de la manière suivante :

Dans le chapitre 2, nous évoquerons les grandes lignes de l'évolution de la reconnaissance des visages. Ce chapitre n'a pas pour finalité de décrire tous les algorithmes de reconnaissance par le visage car cet objectif ne serait pas réaliste, mais nous présenterons les algorithmes qui ont marqué ce domaine et tout au long du chapitre, nous observerons la progression de la recherche dans ce domaine ainsi que l'amélioration des performances. Une deuxième partie de ce chapitre sera consacrée aux solutions apportées aux différents problèmes que rencontre la reconnaissance par le visage, tandis que la dernière partie de ce chapitre présentera l'utilisation des filtres de Gabor dans la biométrie du visage.

Dans le chapitre 3, nous présenterons les outils nécessaires à l'évaluation des performances des algorithmes biométriques. Nous détaillerons les notions de vérification et d'identification ainsi que les bases de données publiques que nous avons utilisées tout au long de la thèse, à savoir : la grande base de visages FRGC (*Face Recognition Grand Challenge*) et la base BANCA. Nous présenterons également une nouvelle base d'évaluation qui complète les précédentes, la base IV².

Dans le chapitre 4, plusieurs algorithmes de reconnaissance, basés sur l'approche par réduction d'espace, sont étudiés. Nous nous focaliserons sur plusieurs problématiques comme le choix des conditions adéquates d'apprentissage ainsi que le choix du meilleur algorithme pour le reste de nos études.

Dans le chapitre 5, nous évoquerons la normalisation géométrique des images des visages mais nous étudierons principalement les différentes normalisations photométriques. Nous proposerons également une nouvelle méthode de normalisation de l'illumination qui donne de meilleurs résultats que les différentes méthodes couramment utilisées dans la littérature, lorsque les conditions d'acquisition sont dégradées.

Dans le chapitre 6, nous introduirons l'utilisation des filtres de Gabor dans la

reconnaissance. Plusieurs travaux font référence à cette approche dans la littérature, en utilisant classiquement l'amplitude. Notre travail principal est de montrer que l'utilisation de la phase des réponses de Gabor apporte une contribution dans l'amélioration des performances. L'utilisation de la phase sera détaillée dans ce chapitre et les résultats comparatifs par rapport à l'utilisation de l'amplitude seule seront donnés dans ce même chapitre. Nous avons testé notre approche sur la base d'évaluation internationale organisée par le NIST, FRGC. Parallèlement à ces évaluations, nous avons testé notre méthode sur une base française provenant du projet national, IV² ("Identification par l'Iris et le Visage via la Vidéo"). Cette méthode a démontré sa robustesse dans les conditions dégradées ainsi que les conditions contrôlées.

Dans le chapitre 7, nous introduirons de nouvelles mesures de qualité des images de visages et nous présenterons différentes stratégies afin d'améliorer les résultats de la reconnaissance. Ces stratégies se basent sur la classification de types d'images suivant les mesures de qualité des images de test. Une introduction de ces mesures de qualité dans le calcul du score de comparaison montrera un apport effectif, en terme de performance, de cette nouvelle approche dans le cadre d'une utilisation d'un modèle de référence et d'un modèle de test utilisant plusieurs images. Une validation sur la base BANCA de cette stratégie sera faite dans la dernière partie de ce chapitre.

Chapitre 2

État de l'art de la reconnaissance de visage

2.1 Introduction

L'identification et/ou la vérification des visages ont attiré l'attention des chercheurs depuis quelques décennies, et restent encore et toujours un sujet de recherche attractif et très ouvert. Beaucoup de connaissances dans les domaines de la reconnaissance des formes, du traitement d'images, des statistiques ont été appliquées au domaine de la reconnaissance du visage. En plus, les capacités grandissantes des moyens informatiques et l'existence de bases de données de grande taille ont permis de mettre au point des algorithmes et des approches de plus en plus complexes et par conséquent, les performances de reconnaissance se sont trouvées améliorées.

Dans les années 90, le but de la recherche dans ce domaine se concentrait sur l'identification des visages enregistrés dans de bonnes conditions, dans des bases de taille relativement réduite. Les protocoles d'évaluation, qui en ont découlé, ont permis de mettre en place les premiers outils d'évaluation et de comparaison des algorithmes. Il faut savoir que les moyens informatiques de l'époque ne permettaient pas d'effectuer des évaluations sur des grandes bases de données.

Dès le milieu des années 90, des campagnes d'évaluations des systèmes de reconnaissance des visages ont été mises en place avec les 3 campagnes d'évaluations FERET (1994, 1995, 1996) [32,37,71]. Les conclusions de ces campagnes sont que

les systèmes proposés étaient encore au stade de développement et que les performances de reconnaissance montraient que le domaine n'était pas encore assez mature. À la fin des années 90, la reconnaissance par le visage connaît un grand essor avec le développement de nouvelles approches et de nouveaux algorithmes, permettant ainsi de passer du domaine de la recherche à la commercialisation. Les premiers systèmes de reconnaissance du visage sont commercialisés dès la fin des années 90 avec les produits des entreprises [Table 2.1]. Dans le même temps, des campagnes d'évaluations à plus grande échelle sont mises en place et les évaluations FRVT (*Face Recognition Vendor Test*) voient le jour, avec FRVT2000 et FRVT2002 [31, 36]. Les conclusions de ces évaluations sont très intéressantes et mettent à jour de nouveaux problèmes qui accompagnent la recherche dans le domaine de la reconnaissance par le visage, à savoir que les performances des systèmes d'identification sont sensibles à la taille des bases de données et que la reconnaissance des visages dans des conditions difficiles a encore besoin d'amélioration.

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

TABLE 2.1: Liste des participants aux campagnes FRVT (source [36])

Les campagnes d'évaluations ont accompagné l'évolution des systèmes de reconnaissance et un besoin d'évaluation statistique plus significative s'est fait ressentir. Par conséquent, des bases de données de visages de plus en plus grandes, avec des conditions de capture plus difficiles et plus complexes ont été acquises. La base de données d'évaluations FERET se composait de ~ 4000 images en 1996, de près de 14000 images en 2000 et d'environ 121 000 images en 2002. Toutes ces

campagnes d'évaluations ont permis à la communauté scientifique d'obtenir des bases de données utiles pour la recherche théorique.

La dernière campagne d'évaluations FRVT2006 a été précédée par la distribution de la base de données FRGC version 1 et version 2, qui reste la plus importante base de visages publique. Cette campagne était différente de la campagne FRVT2002 dans le sens où elle était ouverte à des entreprises ainsi qu'aux laboratoires de recherche. Les organisateurs se sont intéressés dans cette campagne à deux problèmes majeurs dans la vérification des visages, la vérification des visages de face dans les conditions contrôlées et dans les conditions dégradées et la variation des performances suivant la résolution des images de test.

2.2 État de l'art des techniques de reconnaissance

Les avancées applicatives de la reconnaissance par le visage ont suivi rapidement les avancées de la recherche théorique. Nous commençons la description de l'évolution de la recherche par une description de l'objet à reconnaître. Le visage est un objet à nature tridimensionnel. Cet objet peut être sujet à diverses rotations, non seulement planes mais aussi spatiales, et également sujet à des déformations dues aux expressions faciales. La forme et les caractéristiques de cet objet évoluent aussi dans le temps.

Les informations de cet objet tridimensionnel sont capturées à l'aide de caméras et à la sortie, nous obtenons une image bi-dimensionnelle qui le caractérise. Cette réduction de la dimension de l'information permet déjà de faire ressortir quelques problèmes qui se posent lorsque les images 2D ne sont pas parfaitement frontales. D'autres problèmes qui se posent à la reconnaissance des visages sont les conditions environnementales d'acquisition.

Dans l'état de l'art, trois familles d'algorithmes ressortent. Une première famille, la plus populaire, est basée sur une approche globale qui s'appuie sur des méthodes de réduction d'espace. Une deuxième famille d'algorithmes est basée sur des approches locales, appelées aussi "approches par points caractéristiques" et enfin, des algorithmes basés sur des approches hybrides comme l'ACP modulaire (Analyse en Composante Principale) [66], la LFA (*local feature analysis*) et l'analyse des caractéristiques locales [65].

Partant de ces méthodes classiques de reconnaissance, des méthodes plus éla-

borées et plus robustes ont été proposées. Comme exemples d'extension des approches globales, on peut citer les méthodes à base de noyaux [56,57,81], qui ont permis d'améliorer de façon significative les performances de la reconnaissance et dans le cas d'extension des approches locales, la méthode LBP (*local binary pattern*), qui a considérablement amélioré les performances et la rapidité de la reconnaissance [1,27,61].

Dans cette description de l'état de l'art et vu le grand nombre de méthodes appliquées à la reconnaissance du visage, la citation de la totalité des approches et des méthodes ne peut être complète. Je me suis donc intéressé à décrire la plupart des grandes méthodes classiques et la majorité des approches qui ont apporté une vraie originalité et une avancée sensible dans le domaine. Dans la première section, je commencerai par un résumé des algorithmes liés aux méthodes de réduction d'espace ainsi que les résultats qui ont été publiés. Dans la deuxième partie, je résumerai les méthodes les plus populaires des algorithmes basés sur les points caractéristiques.

2.2.1 Méthodes de réduction d'espace

Les méthodes de classification par réduction d'espace restent parmi les techniques les plus utilisées et les plus populaires dans le domaine de la reconnaissance du visage 2D. Ces méthodes considèrent l'image entière du visage comme un vecteur dans un espace multidimensionnel. L'objectif de ces méthodes est de trouver une base réduite de projections qui optimise un certain critère et l'espace original est ensuite projeté sur cette base de dimension inférieure. De ce fait on réduit considérablement la complexité des calculs tout en conservant l'information discriminante entre les images. La classification est habituellement effectuée selon une mesure de distance simple dans l'espace multidimensionnel final.

Divers critères ont été utilisés afin de trouver les bases des espaces de dimensions réduites. Certains d'entre eux ont été définis afin de trouver les bases de projections qui gardent le maximum de variance entre les données, en émettant l'idée que les caractéristiques discriminatoires sont les informations les plus communes entre les données. Le point faible de cette approche est qu'elle n'emploie pas l'information de séparabilité entre les différentes classes.

Une deuxième classe de critères est celle qui traite directement la discrimination entre les classes, et la dernière utilise le critère d'indépendance statistique des données originales dans l'espace de dimension réduite. Une des méthodes les plus

anciennes et les mieux étudiées, pour la représentation des visages dans un espace de dimension réduite, reste l'analyse en composante principale (ACP). Elle appartient à la première classe d'algorithmes [39]. Cette représentation a été appliquée au visage en 1991 par Matthew Turk [93]. L'idée est de trouver une transformation linéaire dans un espace de dimension réduite qui maximise la variance des projections des échantillons originaux. En 1996, l'approche de l'ACP a été étendue à la version non linéaire par l'introduction des fonctions à noyaux non linéaires, appelée "*Kernel Principal Component Analysis*" (KPCA) [81]. L'hypothèse est que si les visages ne sont pas linéairement séparables dans l'espace d'entrée, ils seront linéairement séparables dans un espace non linéaire, d'où l'introduction de notion de noyau de projection non linéaire. Une autre méthode qui vise à représenter le visage sans employer la notion de classe est la factorisation non négative des matrices (*Non Negative Matrix Factorization*, NMF) [46]. L'algorithme de la NMF, comme celui de l'ACP, représente le visage comme combinaison linéaire de vecteurs de la base de l'espace réduit. La différence est que la NMF n'autorise pas les éléments négatifs dans les vecteurs de la base ni dans les poids de la combinaison linéaire. Au final, certains vecteurs de l'espace réduit par l'ACP, appelés "visages propres", ressemblent à des versions distordues du visage entier alors que ceux réduits par la NMF sont des objets localisés qui correspondent mieux à la notion de parties du visage [46].

L'analyse discriminante linéaire (*Linear Discriminant Analysis*, LDA), appelée également "*Fisher Linear Discriminant*", est parmi l'une des méthodes les plus utilisées pour la reconnaissance de visage. Elle utilise le critère de réduction qui se base sur la notion de séparabilité des données par classe. L'idée de la réduction par séparation des classes date de 1936 [21]. Elle a été appliquée aux visages en 1996 [89]. Cette méthode comporte deux étapes : une réduction de l'espace d'origine par l'ACP, puis les vecteurs de l'espace de projection final, appelés "*fisherfaces*", sont calculés sur le critère de séparabilité des classes mais dans l'espace réduit. Ce besoin de réduction de l'espace d'entrée est dû à un critère de singularité de la matrice d'éparpillement totale de l'approche LDA (pour plus de détails, voir le chapitre 4).

Récemment, des algorithmes pour l'application directe de la LDA, pour l'extraction des informations discriminantes, ont été proposés. Ces solutions ont pour but d'éviter la perte d'information discriminante qui se produit après une réduction par ACP. On cite la DLDA (*Direct LDA*) [103] et la RLDA (*Regression*

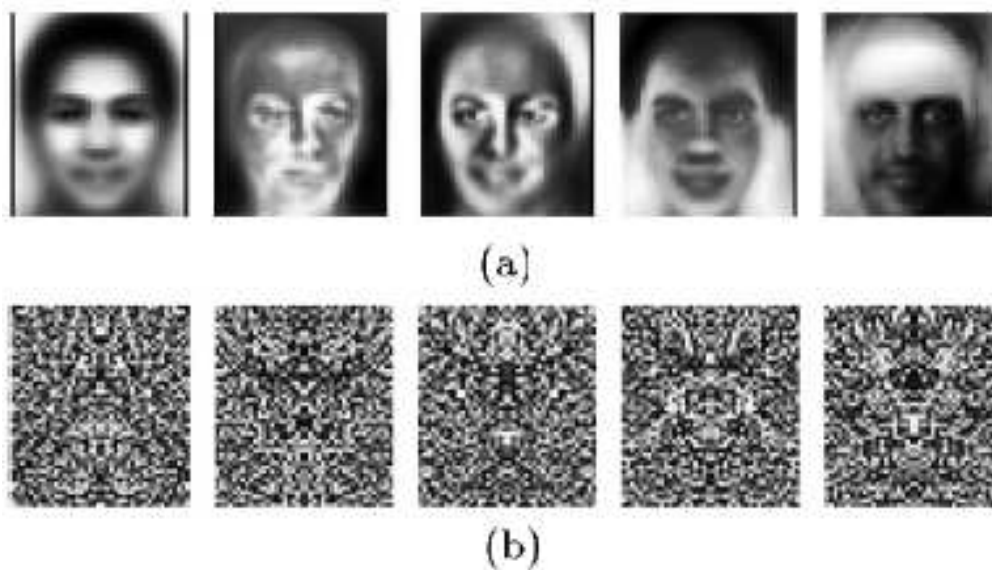


FIGURE 2.1: Exemples des 5 premiers visages propres construits avec l'approche ACP (a) et Exemples des 5 premiers visages propres construits avec l'approche LDA (b)

LDA) [50]. Toutes ces méthodes proposent des approches de diagonalisation directe pour trouver les "*fisherfaces*". Le cadre théorique de la LDA classique, la DLDA et la RLDA, sera étudié en détails dans le chapitre 4.

De même que l'ACP, la LDA a été généralisée pour le cas non linéaire, à savoir l'analyse discriminante généralisée (*General Discriminant Analysis*, GDA) [56] et l'analyse non linéaire discriminante de Fisher (*Kernel Fisher Discriminant Analysis*, KFDA) [57]. L'idée est d'appliquer la LDA à un espace de plus grande dimension (que celui des visages d'entrée), dans lequel les classes des visages sont censées être linéairement séparables. Cet espace est construit par une transformation non linéaire de l'espace d'entrée par des fonctions à noyaux non linéaires. Le cadre théorique de ces deux méthodes et leurs différences seront brièvement présentés dans le chapitre 6.

L'inconvénient principal des méthodes qui emploient des critères discriminants est qu'elles peuvent être sujettes à un surentraînement et la généralisation sur les classes, qui n'appartiennent pas à la base d'apprentissage des vecteurs de projection, peut se trouver affaiblie. D'ailleurs, dans [52], l'auteur montre que dans des conditions d'apprentissage sur des petits ensembles, les vecteurs discriminants n'ont aucune capacité de généralisation et que, dans ce cas, l'utilisation de l'ACP

est plus recommandée. Il a montré, par ailleurs, que la LDA surpasse en termes de performance l'ACP quand de grands ensembles représentatifs de chaque classe de visage sont disponibles.

La dernière classe de critère de réduction, qui a été proposée, est basée sur la réduction d'espace par minimisation de la dépendance statistique des projections, appelée ICA (*Independent Component Analysis*) [6]. Une version non linéaire de cette méthode, appelée KICA (*Kernel Independent Component Analysis*), a aussi été proposée dans [3].

2.2.2 Approches par points caractéristiques

Une autre classe très connue des approches utilisées pour la reconnaissance par le visage est la classe des algorithmes de reconnaissance par points caractéristiques. Les premières méthodes de reconnaissance du visage par ordinateur ont été proposées par Kelly en 1971 [38] et Kanade en 1973. Ces méthodes utilisent la mesure des distances entre les différents points du visage afin de calculer la similarité entre deux visages à comparer. Une autre méthode qui se base sur les points caractéristiques a été proposée en 1994 [78]. Cette méthode utilise les modèles de Markov caché pour l'identification des visages.

La méthode la plus populaire et la plus efficace, dans les approches par points caractéristiques, reste de loin la correspondance élastique des graphes (*Elastic Graph Matching*, EGM). L'idée originale, qui se base sur une architecture de liens dynamiques pour la reconnaissance des objets, a été proposée par Lades en 1993 [44] et a été appliquée, avec succès, à la reconnaissance par le visage, par Wiskott en 1996 [101]. Une variante a également été proposée par le même auteur dans [100]. Cette méthode consiste à construire un graphe de référence en recouvrant l'image du visage référence par une grille et en calculant les réponses des filtres de Gabor à chaque noeud du graphe. Le processus de comparaison est mis en application par une optimisation stochastique d'une fonction de coût, qui tient compte des similitudes des réponses des filtres et de la déformation du graphe. La figure 2.2(a) montre un exemple d'un graphe de visage.

Depuis son application, l'EGM pour la reconnaissance de visage est devenu un champ de recherche très actif. Dans [104], l'auteur a montré que l'EGM surpasse, en termes de performances, l'ACP et des réseaux neurones appliqués à l'identification du visage. Dans [100], l'approche du graphe a été améliorée en introduisant une structure de pile pour chaque noeud. En d'autres termes, chaque noeud

contient des réponses des filtres avec différents états du noeud. Cette approche est appelée "*Elastic Bench graph matching*" (EBGM). La figure 2.2(b) montre un exemple d'un EBGM d'une image de visage. Par exemple, dans le cas de noeuds identifiant les coins de la bouche, les réponses des filtres de Gabor seront calculées sur la bouche ouverte, fermée, souriante..., et ceci pour tous les noeuds.

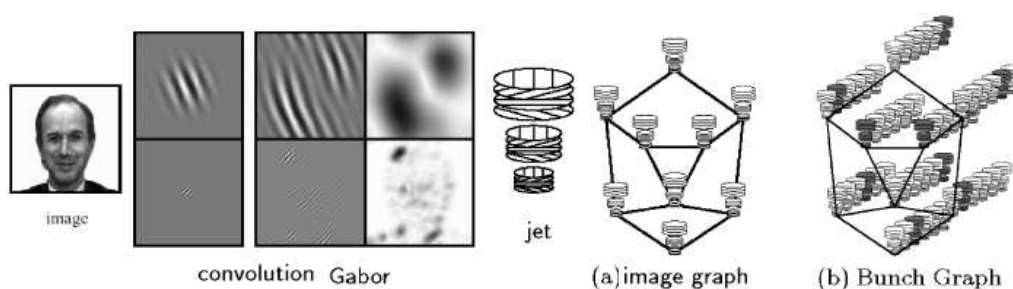


FIGURE 2.2: Structure des approches EGM et EBGM

Dans [99], Wiskott a utilisé l'approche du graphe afin de déterminer certaines caractéristiques du visage tel que la barbe ou les lunettes ou encore le sexe du sujet. Dans le même article, l'auteur a présenté une recherche sur l'influence du coût de l'élasticité des graphes. Afin d'augmenter la robustesse de l'approche par EGM, des solutions pour corriger les problèmes de translations et de déformation ont été proposées dans [102].

Une variante très connue de l'EGM est "la correspondance élastique des graphes morphologiques", appelée "MGEM" (*Morphological Elastic Graph Matching*). L'idée originale de cette variante a été proposée dans [33]. Dans [40, 41], l'auteur a appliqué cette approche pour la reconnaissance des images de visages prises de face dans des conditions d'acquisition diverses. Cette méthode consiste à remplacer les réponses des filtres multi-résolutions de Gabor par des coefficients obtenus par dilatation et par érosion de l'image du visage par un élément structurant [41]. Dans [41], les résultats montrent que l'approche MGEM donne des résultats comparables à ceux obtenus par l'approche classique mais avec l'avantage que cette méthode ne nécessite pas le calcul des banques des coefficients des filtres de Gabor [100].

Une autre méthode se basant sur l'idée des graphes a été présentée dans [92]. La décomposition par approche morphologique a été utilisée comme remplacement de l'analyse par filtres de Gabor et l'auteur a introduit aussi une sélection des pa-

ramètres discriminants par SVM (Séparateur à Vaste Marge, appelé aussi "Machine à Vecteurs de Support" (*Support Vector Machine*)) [75]. D'autres méthodes de sélection ont été étudiées dans [41, 100]. Différents schémas de pondération des noeuds ont été étudiés dans [41, 42, 91, 92].

Comme indiquées dans le paragraphe 2.2.1, les approches globales considèrent le visage comme un vecteur dans un espace multidimensionnel. Le point faible de ces approches est le fait que les images du visage doivent être normalisées avant la phase de mise en correspondance, avec un alignement parfait des différents composants du visage tels que les yeux, le nez et la bouche. Cependant, cet alignement ne peut être parfait qu'avec une bonne détection (automatique ou manuelle) de ces parties. L'avantage avec l'approche EGM et ses variantes est qu'une détection parfaite des parties du visage n'est pas requise. Le point faible de ces méthodes reste le temps de calcul des paramètres et la complexité de la comparaison entre référence et test.

2.3 Influence des variations environnementales sur la reconnaissance par le visage

L'influence de la différence d'illumination entre une image de référence d'un visage et une image de test est bien connue dans le domaine de la reconnaissance faciale. La variation de l'illumination pose beaucoup de problèmes à la majorité des algorithmes de reconnaissance. L'influence de la géométrie du visage et surtout le passage d'une information 3D à une information 2D pose également des problèmes aux chercheurs. Dans le cas d'une comparaison entre un visage de référence et un visage de test avec une orientation spatiale différente, la tâche de reconnaissance devient beaucoup plus difficile. Dans cette section, je présente les solutions les plus connues et les plus efficaces qui ont été présentées dans la littérature.

Pour remédier aux problèmes de l'illumination dans le cadre de la reconnaissance par le visage, trois approches sont proposées dans la littérature : la première s'appuie sur la correction de l'illumination dans l'image, la deuxième sur l'alignement de l'illumination entre l'image de référence et l'image de test, la troisième sur l'élaboration d'algorithmes de reconnaissance robustes aux problèmes de l'illumination.

Le premier algorithme utilisé pour corriger les problèmes de la variation de

l'illumination est l'égalisation d'histogramme. Le point faible de cette méthode réside dans le fait qu'elle ne prend pas en compte les caractéristiques locales du visage. Elle transforme la distribution globale des niveaux du gris d'une forme à une autre sans tenir compte des caractéristiques locales du visage.

En 2003, Wang [97] présente une méthode pour modéliser un visage sous plusieurs conditions d'illumination variables. Il démontre que l'espace de l'illumination d'un visage peut être construit à partir d'au moins 3 images du même visage avec des angles d'illumination non planaires. L'illumination de n'importe quelle image du visage peut être représentée par un point dans cet espace d'illumination. Il montre également que les illuminations des angles extrêmes, appelés aussi "les rayons d'illumination extrêmes", couvrent toute la sphère de l'espace de l'illumination. En pratique, les images avec les rayons extrêmes peuvent ne pas exister. Dans cet article, l'auteur démontre qu'il suffit de quelques images du même visage avec des illuminations diverses pour synthétiser n'importe quelles conditions d'illumination. Lors de cette étude, une méthode d'alignement de l'illumination est également proposée. Cette méthode permet d'appliquer les conditions d'illumination d'une image de visage à une autre image de visage. Ce travail a permis d'introduire le concept d'espace d'illumination. Les résultats expérimentaux ont montré qu'effectivement, la synthèse d'un visage sous différentes conditions d'illumination est possible et que l'alignement de l'illumination améliore sensiblement les performances de reconnaissance. Bien que les auteurs aient montré que la construction d'un espace d'illumination est possible à partir d'un échantillon d'images, la méthode de construction d'un espace d'illumination optimal reste un sujet de recherche. Le fait de construire un espace d'illumination à partir d'exemples de chaque personne ou la construction d'un espace général à partir de plusieurs visages reste aussi un sujet ouvert.

Les travaux, présentés dans [97], ont suivi les travaux de Shashua et Tammy [76,83] qui ont proposé la première approche pour s'affranchir de la luminosité en utilisant l'image quotient. En supposant qu'un visage est un objet à surface lambertienne sans ombres, ils ont montré que le rapport entre l'image de visage test et une combinaison linéaire de trois images du même visage avec des illuminations non-coplanaires aboutit à une image de texture indépendante des variations d'illumination. L'avantage de cette méthode est qu'elle propose une approche simple et robuste pour les algorithmes de reconnaissance. Cependant, cette méthode n'est plus efficace lorsque les visages portent des ombres.

Les travaux sur le cône d'illumination de Belhumeur et Georghiades [7,23,24] ont permis d'expliquer les variations des propriétés du visage suivant les directions de l'illumination. La prise en compte des ombres dans cette approche a aidé à l'amélioration des résultats de la reconnaissance d'une façon significative par rapport aux précédentes approches. Les points faibles sont que cette approche a besoin d'au moins sept images du même visage pour synthétiser le modèle et qu'elle est très coûteuse en temps de calcul.

Terence et al [86] ont proposé une méthode statistique pour modéliser le visage mais cette méthode requière un grand nombre d'images avec des illuminations différentes afin de construire l'espace d'illumination des visages.

Pour s'affranchir des problèmes d'illumination, les auteurs dans [73] ont proposé une nouvelle méthode pour ajuster l'illumination des images de visages vers une illumination standard. Cet ajustement est basé sur la ressemblance des visages. Cette approche se base sur la constatation que la forme de la surface des visages humains est quasiment identique. Dans l'algorithme, une estimation des 9 plus basses composantes fréquentielles de l'illumination est faite, ensuite l'illumination du visage est réajustée à une illumination standard en se basant sur l'image quotient [76]. Dans le cas de la reconnaissance des visages, deux illuminations standards ont été proposées : une illumination uniforme et une illumination ponctuelle frontale. La première illumination réajuste l'information de la texture et la deuxième réajuste non seulement celle de la texture mais aussi l'information des zones d'ombre. Les résultats expérimentaux ont montré que l'approche proposée améliore sensiblement les résultats et que les performances de l'illumination uniforme sont un peu mieux que celle de l'ajustement avec une illumination ponctuelle frontale. Ce résultat, inattendu par les auteurs, s'explique par le fait que l'illumination ponctuelle frontale a besoin d'une labélisation précise entre les zones de texture et les zones d'ombre. Cela peut ne pas être le cas pour un système de labélisation automatique. Les résultats de cette approche se dégradent beaucoup dans les conditions d'illumination extrêmes.

Dans [74], le même auteur a proposé une méthode de reconnaissance basée sur la dérivée de l'image logarithmique (obtenue par le gradient de la transformée logarithmique de l'image originale) . Il a aussi décrit une mesure de distance qui se base sur le minimum entre les dérivées de deux images à comparer. Les résultats de cette distance , appliquée à la base de visage CMU-PIE [85], montrent que cette méthode est beaucoup plus performante que la corrélation de l'image d'intensité

et qu'elle donne de meilleurs résultats que la simple distance euclidienne entre les dérivées des images logarithmiques.

La normalisation de l'illumination a été très bien explorée dans la littérature. Différentes approches pour compenser les variations de l'illumination pour les visages ont été décrites. Une des premières constatations est que la normalisation de l'illumination peut dégrader les résultats lorsque les visages dans la base sont illuminés de façon uniforme. En effet, dans ce cas, une normalisation de l'illumination peut supprimer l'information discriminante entre les visages ; Ceci est aussi vrai dans le cas où la base de données contient des visages de personnes avec des couleurs de peau différentes. Concernant les autres variabilités qui peuvent affecter la reconnaissance par le visage (pose, expressions ...) , je ne cite ici que quelques références, sachant que c'est un domaine où la recherche est foisonnante et où les problèmes sont encore mal résolus

En 2004, Chang et Bowyer [14] ont présenté leurs travaux sur la combinaison de l'image 3D et de l'image 2D du visage. Cette combinaison de l'image de profondeur et de l'image de texture a montré une amélioration significative par rapport à l'utilisation de chaque modalité toute seule. L'apport de ce travail s'appuie sur l'utilisation d'une base de données de grande taille et sur l'existence d'une grande différence temporelle entre les sessions d'acquisitions (~ 13 semaines en moyenne entre les données de référence et les données de test). Les mêmes auteurs ont publié, en 2005 [15], leurs travaux en utilisant des espaces de réduction indépendants entre la 2D et la 3D. Leurs conclusions sont les suivantes :

- Les résultats de performances pour la 2D et la 3D sont équivalents
- Une fusion 2D+3D améliore les résultats de façon significative par rapport à la 3D ou à la 2D utilisée seule.
- La fusion des scores de plusieurs images 2D par modèle améliore les résultats par rapport à l'utilisation d'une seule image 2D par modèle.
- Même en utilisant plusieurs images par modèle, la reconnaissance en utilisant la fusion 2D+3D reste meilleure. Cependant, l'amélioration des performances devient moins significative lorsqu'on n'utilise qu'une seule image.

Les résultats présentés utilisent la méthode de réduction d'espace PCA pour les deux modalités fusionnées (2D et 3D). Les auteurs précisent qu'une utilisation d'algorithmes plus adaptés pour chacune des modalités peut améliorer les performances d'une modalité par rapport à l'autre. Selon leur conclusion, l'utilisation de plusieurs images du même visage, sous différentes conditions d'illumination

et d'expression, reste la solution la plus fiable pour améliorer la reconnaissance. Dans cette approche, le problème de pose n'est pas pris en compte mais les auteurs précisent qu'il faut traiter le problème de l'illumination et de la pose simultanément. La méthode, exploitant les modèles multi-images pour représenter une personne, est peu explorée dans le domaine de la reconnaissance par le visage.

Dans [28], les auteurs proposent une autre méthode pour palier le problème de variation de la position du visage. Leur algorithme, basé sur la transformation discrète par cosinus (DCT, *Discrete Cosine Transform*), reste robuste face aux variations de pose, de taille et d'alignement des visages. Le système avait été testé sur différentes bases de données de référence. Cet algorithme utilise une transformation affine des coefficients de la DCT afin de corriger ces problèmes géométriques.

Un des problèmes majeurs qui influence la reconnaissance par le visage reste l'expression faciale. On s'intéressera, par la suite, à la description de deux étapes importantes qui sont l'extraction des caractéristiques de l'expression et la reconnaissance et la classification de cette expression faciale suivant des règles de classifications.

La majorité des travaux de détection et de classification des expressions, décrits dans la littérature, se basent sur les réseaux de neurones. Ces réseaux sont appliqués soit directement à l'image du visage, soit à l'image réduite par l'ACP ou par l'ICA ou soit à l'image convoluée par des filtres de Gabor. Leur but est de classifier les différentes expressions.

Dans [20], Fasel a développé un système basé sur des réseaux de neurones à convolution afin d'augmenter la robustesse face aux problèmes de translation et de taille. Dans [17], l'auteur a proposé une méthode basée sur les réseaux de neurones. La méthode s'applique en 3 étapes : une extraction des caractéristiques est effectuée en appliquant un filtrage de Gabor sur des points fixes du visage, puis une réduction de la dimension par l'ACP est entreprise sur les réponses des filtres et enfin, les données réduites sont appliquées à l'entrée d'un réseau de neurones à six sorties, correspondant aux 6 expressions faciales.

Dans [82], une autre approche a été exploitée. Elle se base sur la machine à support de vecteur (SVM) pour classifier les différentes expressions.

Des méthodes d'apparence et des méthodes basées sur la géométrie du visage ont aussi été proposées. Pour les méthodes d'apparence, des points de référence sont sélectionnés manuellement ou automatiquement. Des filtres de Gabor sont

appliqués à ces points et les réponses de ces filtres formeront un vecteur de caractéristique, une classification de ce vecteur est ensuite faite. Le filtrage de Gabor peut aussi être appliqué sur toute l'image au lieu des points de référence.

Pour les méthodes basées sur la géométrie des visages, les positions de points de référence sont utilisées afin de classifier les différentes expressions. Les méthodes d'apparences semblent donner des résultats de classification satisfaisants. La combinaison des deux approches, présentée dans [47], montre de bien meilleures performances que la méthode d'apparence seule, basée sur les ondelettes de Gabor.

Une technique de référence très utilisée pour la classification des expressions est basée sur la discrimination des coefficients de projection sur l'espace ICA soit par un classifieur de similarité cosinus (appelé aussi "mesure cosinus") ou soit par un classifieur se basant sur le maximum de corrélation.

Deux systèmes hybrides pour la classification de sept expressions faciales ont été proposés dans [30]. Le premier système combine l'ICA et le SVM. L'image de visage est projetée sur l'espace construit par l'ICA et le vecteur des coefficients de projection est ensuite appliqué à l'entrée du SVM. Les performances du système sont meilleures que celles du système de référence. Le deuxième système est basé sur la convolution de l'image du visage par des filtres de Gabor et la nouvelle représentation du visage est utilisée comme vecteur d'entrée pour le classifieur. Trois classifieurs ont été étudiés : le SVM, la similarité cosinus et le maximum de corrélation. Les résultats ont montré que la classification par le SVM est meilleure que les deux autres approches.

2.4 Les approches utilisant le filtrage de Gabor

Étant donné qu'une partie de mon travail consiste à proposer une nouvelle méthode globale de vérification par le visage reposant sur l'extraction des caractéristiques par filtrage de Gabor, je cite ici l'état de l'art dans ce domaine .

En 1993, Martin Lades [44] a proposé pour la première fois l'utilisation des filtres de Gabor comme outils d'identification de visage. Il a décrit une approche neuronale basée sur la réponse d'une famille de filtres de Gabor (la première version du EGM).

La raison d'employer seulement l'amplitude est de fournir une mesure monotone des propriétés de l'image. De nombreux travaux ont employé les ondelettes de Gabor. Par exemple, Winscott [101] les utilise dans la comparaison élastique

des graphes (EGM), Ayinde [2] dans la corrélation des représentations des filtres de Gabor. D'autres travaux emploient les approches basées sur la réduction de l'espace comme l'ACP, la LDA, la GDA, la KPCA sur les représentations de Gabor du visage, en exploitant l'amplitude et la partie réelle de la réponse des filtres. En 2004, Liu a employé la KPCA avec un noyau polynomial à puissance partielle appliquée aux représentations de Gabor [48]. Dans [49], le même auteur a utilisé l'approche de réduction d'espace à noyau non linéaire (KFA (*Kernel Fisher Analysis*)), appliquée aux représentations de Gabor. Cela a permis d'améliorer les résultats de l'évaluation FRGCv2 du visage 2D de 12% VR (taux de vérification, ou "*Verification Rate*") à 0.1% de FAR (taux de fausses acceptations, ou "*False acceptance Rate*") du système de référence à 78% VR à 0.1%. C'était de loin l'amélioration la plus importante publiée pour cette base de données.

Dans [84], de nombreuses méthodes de réduction à base de noyaux non linéaires (GDA, KPCA ...) ont été étudiées et comparées aux réductions d'espace classiques.

L'extraction de l'information par les filtres de Gabor a été montrée dans toutes ces publications comme très pertinente au vu des performances obtenues.

2.5 Conclusion

Dans ce chapitre, nous avons présenté les grandes lignes de l'évolution de la reconnaissance par les visages. Comme déjà évoqué dans l'introduction, ce chapitre n'a pas pour finalité de décrire tous les algorithmes de reconnaissance faciale mais nous nous sommes contentés de présenter les algorithmes qui ont introduit une nette évolution dans le domaine de cette biométrie, permettant ainsi une véritable amélioration des performances.

Une deuxième partie de ce chapitre a été consacrée aux solutions apportées aux problèmes spécifiques rencontrés dans ce domaine de recherche, comme les problèmes de variabilité d'expression, d'illumination ou bien de pose.

Au vu des résultats performants obtenus par la modélisation des visages par analyse de Gabor multirésolution, nous avons choisi d'utiliser cet outil d'analyse associé à une méthode de projection linéaire afin de construire notre méthode de reconnaissance.

Chapitre 3

Outils d'évaluation, bases de données et protocoles

3.1 Introduction

Comme déjà présenté dans le chapitre précédent, l'engouement pour les systèmes biométriques a connu un grand essor au début des années 2000. Les dépôts de brevet dans ce domaine se sont multipliés et des systèmes applicatifs à grande échelle ont été mis en place, comme le système utilisé lors des élections présidentielles au Venezuela en 2004 ou le système US Visit, devenu opérationnel depuis 2004. Ces systèmes restent basés sur les empreintes digitales, et ceci est dû aux évolutions techniques des algorithmes pour la biométrie basée sur les empreintes digitales.

Devant une telle expansion des applications, les outils d'évaluation de ces systèmes se devaient alors de suivre cette expansion afin d'assurer une bonne généralisation des performances. Dans l'absolu, pour qu'une évaluation d'un système biométrique soit efficace, elle doit prendre en compte plusieurs critères, qui sont non seulement des critères techniques mais aussi des critères opérationnels. Les performances et la rapidité d'exécution restent les deux critères techniques les plus importants pour une application biométrique. La facilité d'utilisation et la convivialité sont des critères opérationnels qui commencent à avoir leur place dans les évaluations. Par exemple, dans le cas d'un système à base d'empreintes digitales, il devra avoir les meilleures performances de reconnaissance possibles

aussi bien sur des empreintes normales que sur des empreintes abîmées ou peu texturées. Il devra aussi détecter les faux doigts tout en donnant une réponse rapide et en restant simple à l'utilisation.

Comme indiqués dans le chapitre 2, plusieurs efforts d'évaluation et de standardisation des algorithmes basés sur le visage ont vu le jour dès le début des années 90. Ils ont surtout été menés par le gouvernement américain à travers l'organisme du NIST (*National Institute of Standards Technology*) avec les campagnes FERET. Des campagnes d'évaluations sont régulièrement organisées sur le visage [70], la voix [59], les empreintes digitales [60] et l'iris [70]. Toutes ces campagnes entre autres ont proposé différents outils de mesure des performances algorithmiques des systèmes.

La différence entre la biométrie et les techniques classiques de vérification de l'identité réside dans le fait qu'en biométrie, il est impossible d'obtenir une correspondance parfaite entre une représentation de référence et une représentation acquise au moment du test. Ceci est dû aux variations intrinsèques ou extrinsèques des représentations. Pour les autres techniques de vérification telles que l'utilisation de code confidentiel, la correspondance doit être parfaite.

En biométrie, une représentation d'une identité, aussi appelée "gabarit", est un ensemble de caractéristiques extraites de la donnée biométrique (empreinte, visage, iris, voix...). Ces caractéristiques sont sujettes à des variations entre le moment de la capture de la référence et la phase de test. Ces variations sont dues à plusieurs facteurs, tels que des facteurs temporels ou environnementaux. Par exemple, pour les données biométriques comme le visage, l'âge, la prise de poids et la coiffure jouent un rôle dans cette dissimilarité entre deux moments de capture. L'empreinte digitale est aussi une biométrie qui peut être sujette à des variations. Une coupure au doigt ou même des maladies peuvent influencer la correspondance entre les deux représentations. A ces variations intrinsèques des échantillons, nous pouvons ajouter l'influence de l'acquisition et les erreurs de capture. Un capteur ne peut pas restituer une acquisition parfaite d'un même échantillon biométrique, ce qui introduit encore plus de variation pour le même échantillon. Par exemple, pour le visage, les conditions de luminosité ou le positionnement par rapport à la caméra, ou même le changement des capteurs entre les deux phases d'acquisition (référence et test) sont des exemples de facteurs environnementaux. Pour l'em-

preinte, le vieillissement et le salissement des capteurs peuvent également induire des variations entre le moment de la capture de l’empreinte référence et la phase de test. Ce type d’erreurs est dû généralement à la nature même du système biométrique. Cependant, d’autres types d’erreur existent, correspondant à des tentatives d’impostures ou à des tentatives de contournement de la vérification. Un exemple de tentatives d’imposture est l’utilisation de copie d’empreintes et un exemple de tentatives de contournement est le changement d’apparence pour le visage (ajout de barbe ou port de lunettes...).

Tous ces facteurs sont cités pour expliquer qu’en pratique, une correspondance parfaite entre les deux représentations (référence et test) pour le même échantillon biométrique n’est pas possible et qu’un système biométrique peut même être induit en erreur par des tentatives de contournements et d’impostures.

En conclusion, en biométrie, les différences, qui peuvent exister entre la représentation du signal de référence et celle du signal de test, engendrent forcément des erreurs du système biométrique. Par la suite, nous listerons les différentes méthodologies afin de comparer les différentes modalités entre elles, notamment en termes de taux d’erreur, ce qui n’est pas la seule mesure de performance à considérer.

En biométrie, nous utilisons le terme "système de reconnaissance" pour désigner, sans distinction particulière, un système biométrique de vérification ou un système biométrique d’identification. Il faut savoir que les systèmes biométriques peuvent être classés en deux grandes familles :

Le système de vérification d’identité : Dans ce cas, le système biométrique vérifie que l’identité proclamée par la personne devant le système est bien la bonne. Ce mode de test est appelé un test en "un contre un".

Le système d’identification : Dans ce cas, le système doit trouver dans une base de données l’identité de la personne qui se présente devant lui sans aucune connaissance *a priori* sur son identité. Ce mode de test est appelé un test "un contre N". En mode d’identification, plusieurs modes de fonctionnement sont possibles. Nous les détaillerons plus tard dans ce chapitre.

3.2 Scores et protocoles d'évaluation des systèmes biométriques

Par la suite, nous ne citerons pas la totalité des outils d'évaluation d'un système biométrique car la liste serait très exhaustive. En revanche, nous présenterons les outils les plus utilisés par la communauté ainsi que ceux utilisés lors de cette thèse.

Le résultat de comparaison entre deux représentants biométriques par un système de reconnaissance est appelé "score". Ce "score" peut être une mesure de similarité ou une mesure de distance (ou dissimilarité) entre les échantillons comparés. Plus la valeur du score de similarité est grande, plus la probabilité que les deux représentants biométriques appartiennent à la même identité est forte. Respectivement, plus cette mesure est faible, plus la probabilité que les deux représentants biométriques proviennent de deux identités différentes est faible. Inversement, plus la mesure de distance entre deux échantillons est faible, plus la probabilité qu'ils appartiennent à la même identité est grande et vice versa. En général, la mesure de similarité est utilisée pour avoir une indication sur la corrélation entre les échantillons alors que la mesure de distance est plus souvent utilisée pour mesurer le degré d'éloignement entre l'échantillon de référence et l'échantillon de test.

Le but derrière le calcul de ce score est de pouvoir assigner une identité à une personne (identification) ou confirmer ou infirmer l'identité de celle-ci (vérification) en prenant une décision par rapport à la mesure calculée. Cette décision est prise en fixant un seuil, dit "de décision", par rapport au score calculé. Plus concrètement, si $S1$ et $S2$ sont les deux signaux caractéristiques à comparer, si θ est le seuil de décision et si $d(S1, S2)$ est inférieur à θ , alors $S1$ et $S2$ sont considérés comme un signal de la même personne. Sinon, ils sont considérés comme deux signaux différents. Généralement, les évaluations dans le domaine de la biométrie sont basées sur deux types de score. Nous parlons de scores "intra-classes" et de scores "inter-classes". Les scores "intra-classes", comme leur nom l'indique, résultent de la comparaison d'échantillons provenant de la même classe (personne). Les scores "inter-classes" résultent eux de comparaisons d'échantillons biométriques de personnes différentes.

En mettant en place ce que nous appelons, communément, des protocoles d'évaluation, nous obtenons des distributions pour chaque type de score. Les pro-

toques d'évaluation sont, en réalité, des listes de comparaisons "intra-classes" et "inter-classes". Plus le nombre de classes, le nombre d'échantillons par classe et le nombre de tests de comparaison est grand, plus l'évaluation est considérée statistiquement significative. Dans le cas d'un nombre suffisant de classes, d'échantillons et de tests, les distributions des deux types de score convergent vers des distributions normales. Ceci s'explique par le "Théorème Central Limite". La figure 3.1 représente une distribution théorique des scores "intra-classes" et "inter-classes" pour un système biométrique théoriquement parfait. Il est dit parfait dans le sens où si nous choisissons un seuil de décision situé dans la zone "A", aucune erreur de classification n'est faite.

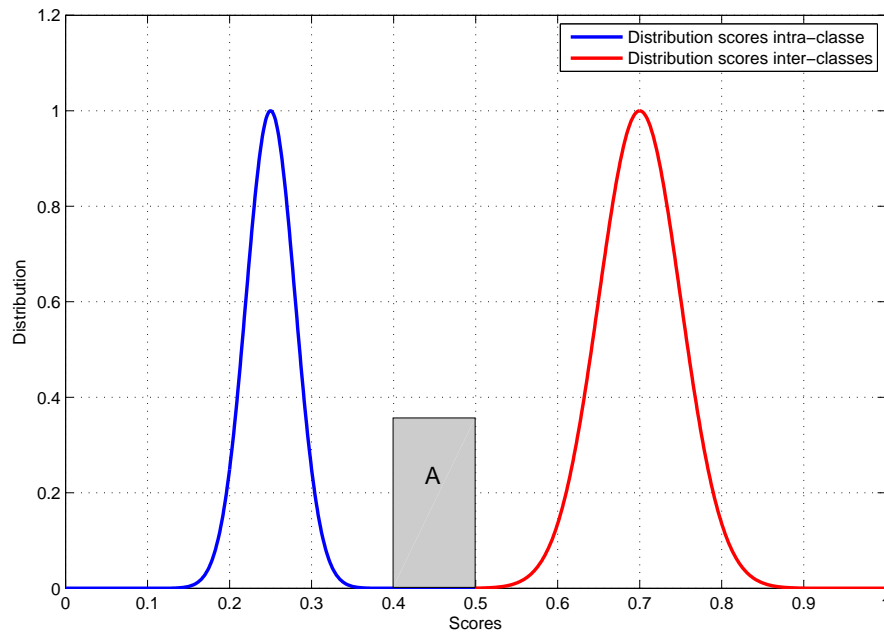


FIGURE 3.1: Exemple de distributions théoriques des scores "intra-classes" et "inter-classes" pour un système biométrique théoriquement parfait

Dans la plupart des applications biométriques et dans le cas où on a un nombre assez élevé de test de comparaisons différentes, les deux distributions "intra-classes" et "inter-classes" se chevauchent. La figure 3.2 présente cette configuration et la zone d'intersection "B" des deux distributions est considérée comme la zone critique de la reconnaissance.

En effet, si on considère un seuil de décision à l'intérieur de cette zone, il existera toujours des scores qui seront du mauvais côté de la frontière ainsi définie.

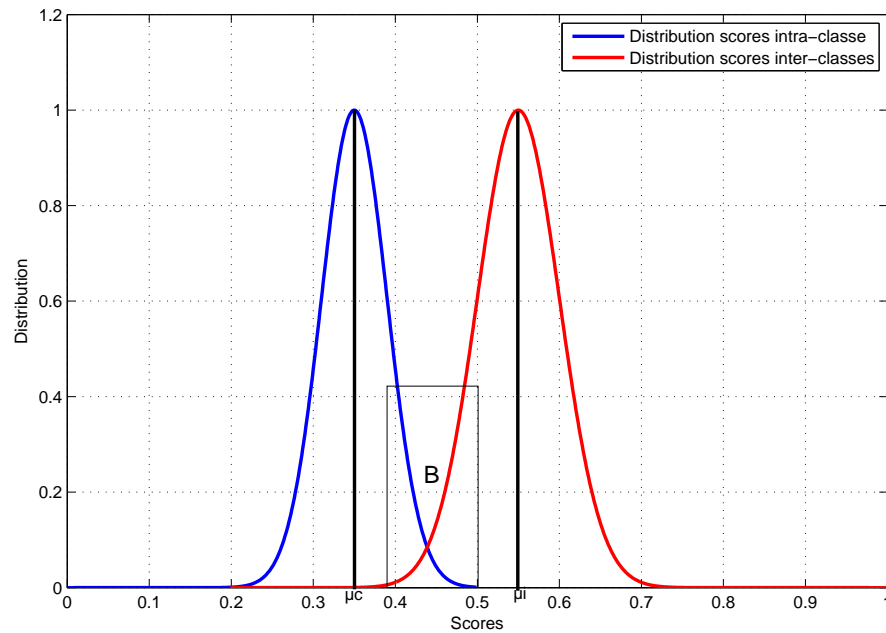


FIGURE 3.2: Exemple de distributions théoriques des scores "intra-classes" et "inter-classes" pour un système biométrique réel

En conclusion, pour qu'un système biométrique soit considéré comme efficace, il faut que le chevauchement entre les scores "intra-classes" et "inter-classes" soit minimal.

3.3 Évaluation des systèmes de reconnaissance

Les performances des systèmes biométriques sont évaluées différemment selon que l'on considère un mode de fonctionnement en vérification ou en identification.

3.3.1 Évaluation des systèmes de vérification biométrique

Deux types d'erreurs peuvent survenir dans un système de vérification d'identité, le faux rejet et la fausse acceptation. Le premier type a lieu lorsque le résultat de comparaison des échantillons d'une même personne est considéré par le système comme un résultat d'une comparaison "inter-classes". La fausse acceptation se produit lorsque le système considère un test d'imposture comme test "intra-classes".

En cumulant sur une base de données et avec un protocole d'évaluation spécifique (définition des accès clients et imposteurs pour chaque personne) les différentes erreurs de décision commises par le système de vérification, nous obtenons les taux d'erreur correspondant à ces deux types d'erreur, une fois un seuil de décision fixé. Le taux de faux rejets, dénommé FRR (*False Rejection Rate*) ou FNMR (*False Non Match Rate*) ou "False Alarm Rate", est le pourcentage des données de test authentiques qui ont été rejetées par rapport au nombre total de comparaisons "intra-classes". Le taux de fausses acceptations, dénommé FAR (*False Acceptance Rate*) ou FMR (*False Match Rate*) ou "Miss Probability", est le pourcentage des comparaisons d'impostures qui ont été acceptées par rapport au nombre total de comparaisons "inter-classes".

On considère un système de vérification produisant une mesure de distance entre deux échantillons biométriques. Soit S_{ij} , le résultat de comparaison de ces deux échantillons pour deux identités i et j . Les taux de FRR et FAR pour un seuil θ donné sont calculés de la manière suivante :

$$\text{FRR}(\theta) = \frac{\text{card}(\{S_{ij}/S_{ij} > \theta\})}{\text{card}\{S_{ij}\}} \text{ avec } i=j \quad (3.1)$$

$$\text{FAR}(\theta) = \frac{\text{card}(\{S_{ij}/S_{ij} < \theta\})}{\text{card}\{S_{ij}\}} \text{ avec } i \neq j \quad (3.2)$$

avec *card* le cardinal de l'ensemble ou le nombre d'éléments de l'ensemble.

Un autre taux d'erreur peut être considéré, c'est le taux d'erreur total TER (*Total Error Rate*). Ce taux est le pourcentage d'erreur totale que commet le système par rapport à tous les tests de comparaison. En réalité, c'est une fusion des deux taux FAR et FRR, il est calculé par :

$$\text{TER}(s) = \frac{\text{card}(\{S_{ij}/S_{ij} > \theta \text{ avec } i=j; S_{ij}/S_{ij} < \theta \text{ avec } i \neq j\})}{\text{card}\{S_{ij}\}} \quad (3.3)$$

Le problème majeur de cette formulation du TER réside dans le fait que dans une évaluation biométrique réelle, le nombre d'échantillons par personne est généralement faible. Ce faible nombre d'échantillons impose un nombre limité de comparaisons intra-classes par rapport au nombre de comparaisons "inter-classes" ou impostures qui sont en général générés par la comparaison des échantillons de différentes personnes de la base. Cette situation conduit alors une variation asymptotique du TER par rapport au taux d'erreur FAR.

Pour éviter ce problème, on considère, dans la pratique, une autre expression du

TER qui est moins sensible à ce phénomène :

$$\text{TER}(\theta) = \frac{\text{FAR}(\theta) + \text{FRR}(\theta)}{2} \quad (3.4)$$

Cette nouvelle formulation est indépendante du rapport relatif entre le nombre de tests intra-classes et celui des tests inter-classes.

Dans tout système biométrique, la variation des taux d'erreur FAR et FRR en fonction du seuil est inversée. On se place toujours dans le cas d'un système qui produit une distance comme résultat de comparaison. Plus le seuil de décision augmente, plus le taux de FAR diminue et celui du FRR augmente. La figure 3.3 donne un exemple de ces variations.

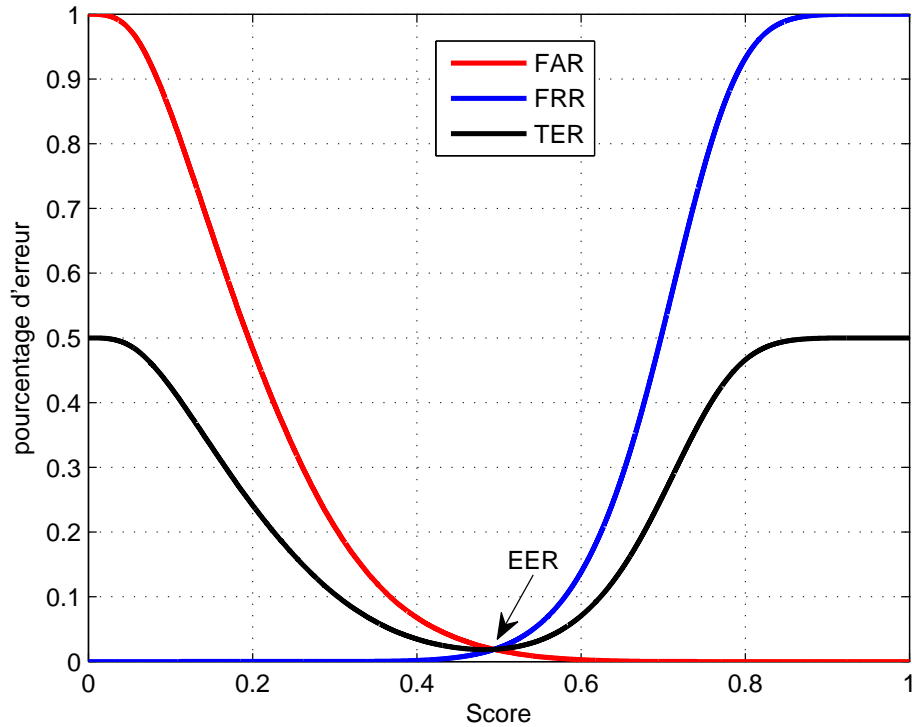


FIGURE 3.3: Variation des taux FAR, FRR et TER en fonction du seuil de décision

On peut remarquer que dans les zones où le seuil est faible, le taux de FAR est grand et celui du faux rejet est faible et inversement, dans les zones où la valeur du seuil est grande. Pour chaque valeur du seuil, une valeur particulière de FAR et de FRR sont associés. Un choix particulier d'un seuil de décision

impose un fonctionnement particulier du système biométrique. Plusieurs points de fonctionnement particuliers sont utilisés pour comparer les systèmes biométriques entre eux :

- EER (*Equal Error Rate*) : c'est le point de fonctionnement historique de la biométrie. Il correspond à un choix de seuil où le taux de FAR est égal au taux de FFR. Ce point est celui de l'intersection entre les deux courbes dans la figure 3.3. Dans la littérature, on parle souvent d'un système biométrique à X% de EER, pour dire qu'il est performant ou pas. Mais on omet souvent de citer la base de données sur laquelle a été testé le système ou le protocole d'évaluation, ce qui rend les comparaisons entre les systèmes impossibles. Ce point de fonctionnement est souvent utilisé comme point de comparaison dans les compétitions biométriques internationales.
- Min HTER ou HTER (*Half Total Error Rate*) : c'est le point particulier où l'erreur totale du système est minimale. Ceci correspond à un minimum global dans la courbe du TER de la figure 3.3. Si les variations des taux FAR et FRR sont comparables, la valeur du HTER sera très proche de celle du EER.
- VR @ 0.1% de FAR : le taux de vérification (équivalent à 1-FRR) est calculé pour un FAR = 0.1%. Ce point de fonctionnement est très utilisé lorsque le nombre de comparaisons pour les tests "inter-classes" (ou "tests d'imposture") est très grand, \gg à 1000, afin qu'il y ait au moins une erreur pour obtenir 0.1%. Ce point est important pour étudier le comportement des systèmes à faible taux de FAR (sécurité accrue) sur une grande base de données ou pour simuler une application à grande échelle. Ce point de fonctionnement était exclusivement utilisé pour les empreintes digitales, mais en raison des améliorations des systèmes de reconnaissance du visage et la mise en place de bases de données très grandes, il est devenu la nouvelle norme pour cette biométrie.

Comparer plusieurs systèmes entre eux, en s'appuyant sur les résultats d'un seul point de fonctionnement, ne peut en aucun cas caractériser le fonctionnement de ce système ou donner une idée sur le meilleur système lors d'une évaluation.

Pour remédier à ce problème, une analyse plus globale est faite et des courbes de performance ont été proposées.

Courbes de performances

On utilise, le plus souvent, les courbes ROC (*Receiver Operating Characteristics* ou *Relative Operating Characteristic*) qui représentent la variation du taux de vérification ($100 - \text{FAR}(\%)$) en fonction du taux de FRR(%). La figure 3.4 montre des courbes ROC de trois systèmes de vérification du locuteur (données NIST). Les courbes ROC doivent être les plus asymptotiques possibles aux axes des abscisses et ordonnées. Dans l'exemple montré dans la figure 3.4, l'algorithme représenté en violet (\times)¹ surpasse ceux représentés en noir (\bullet)¹ et en vert (\blacksquare)¹, qui semblent quant à eux avoir des performances similaires.

Les courbes ROC ont elles aussi leurs limites. Comme montré dans la figure 3.4, la courbe ROC représente la variation du taux de bonne reconnaissance uniformément, quelque soit le taux de FRR. Or, ce comportement devient asymptotique à partir d'un certain taux de FRR qu'il soit très faible ou très fort. Nous pouvons d'ailleurs facilement noter d'après la figure que seul 10% du graphe est utile et informatif. Pour pallier ce problème, on utilise une autre courbe, appelée "courbe DET" (*Detection Error Trade-off*) [51] où les valeurs FAR et FRR sont normalisées de façon non linéaire (logarithmique) afin de permettre à la courbe de s'adapter au fonctionnement du système. La figure 3.5 montre un exemple de courbes DET correspondant aux mêmes systèmes et aux mêmes données que ceux qui ont permis de générer les courbes ROC montrées à la figure 3.5. Généralement, la fonction non linéaire de normalisation des axes utilisée est le logarithme népérien. Avec cette normalisation, dans la majorité des cas, la courbe DET d'un système est presque linéaire, surtout pour les modalités biométriques pour lesquelles les distributions des taux d'erreurs FAR et FRR sont normales. Plus la courbe DET est proche de l'origine, meilleur est le système considéré.

Dans la figure 3.5, nous pouvons remarquer un avantage des courbes DET par rapport aux courbes ROC. En effet, bien que lors de l'analyse des courbes ROC, les systèmes en noir (\bullet) et en vert (\blacksquare) semblaient avoir un comportement quasi-identique, l'analyse des courbes DET montre qu'à faible taux de FRR, le système en noir (\bullet) est nettement supérieur à celui représenté en vert (\blacksquare) .

1. Pour la copie en noir et blanc du manuscrit

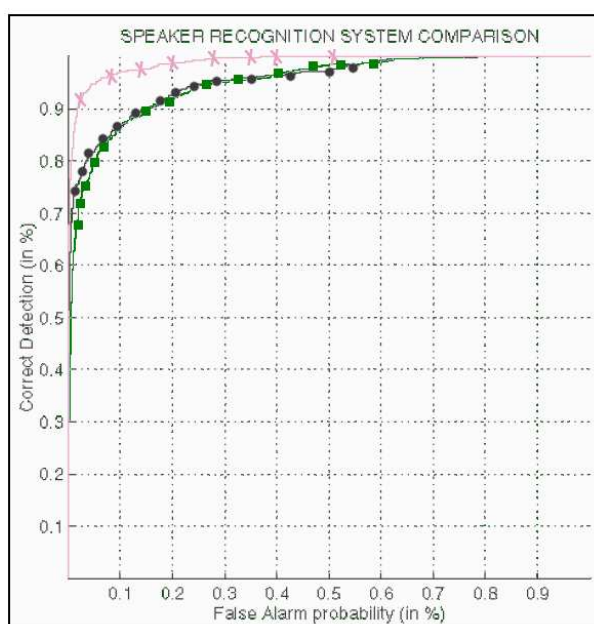


FIGURE 3.4: Exemple de courbes ROC où le taux de vérification correcte est tracé en fonction du taux de FRR (Source NIST reconnaissance de la parole) en échelle linéaire

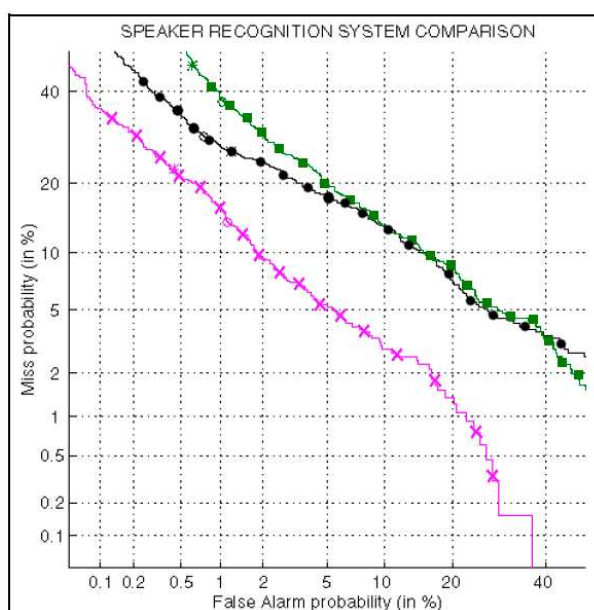


FIGURE 3.5: Exemple de courbes DET où le taux de FAR est tracé en fonction du taux de FRR (Source NIST reconnaissance de la parole) en échelle logarithmique

Intervalle de confiance

Outre le choix des critères d'évaluation des performances des systèmes biométriques, nous devons faire face à une autre difficulté de l'évaluation d'un système biométrique. En général, on ne dispose que de peu de données de chaque personne et de peu de personnes dans les bases. Certaines bases sont composées de quelques dizaines de personnes seulement et on a peu de données par personne, même si récemment des bases de l'ordre de centaines de personnes ont fait leur apparition (par exemple FRGC, MBGC). Il est alors impératif de mesurer en plus des taux d'erreurs calculés, une mesure de confiance associée.

Deux types de mesure des intervalles de confiance sont introduits dans [10] : une basée sur une méthode non-paramétrique appelée "*bootstrap*" et une basée sur une méthode paramétrique. S'ajoutant à ces deux méthodes, la méthode classique du "*leave one out*" n'est pas adaptée aux protocoles d'évaluation biométrique, où les bases d'apprentissage et les listes des tests sont fixées à l'avance (type BANCA, FRGC, XM2VTS...). En effet, cette méthode considère plusieurs découpages des bases d'apprentissage et de tests.

méthode non-paramétrique Pour utiliser la méthode d'estimation d'intervalles de confiance par le "*bootstrap*", il faut réaliser un ré-échantillonnage de l'ensemble de test. Ce ré-échantillonnage consiste en un tirage aléatoire avec remplacement de M exemples. Dans notre cas, on prendra M égal au nombre d'exemples. Pour l'estimation du FRR, on prendra donc $M = N_c$ avec N_c le nombre de scores "client" et pour le FAR, on prendra $M = N_i$, N_i étant le nombre de scores "imposteur".

Pour chaque tirage, on va donc choisir M exemples avec remplacement, c'est à dire que les exemples peuvent être utilisés plusieurs fois ou jamais. Pour chaque tirage, on estime ensuite les valeurs du FRR et du FAR pour une valeur du seuil θ . On répète ensuite cette procédure T fois avec T tirages (en général, 1000 tirages sont suffisants). A partir de ces T valeurs de $FAR(\theta)$ et de ces T valeurs de $FRR(\theta)$, on peut estimer alors l'intervalle de confiance à 90% ou à 95% autour de la médiane. Le problème de cette méthode reste le même que celui du "*leave one out*" à savoir que pour les nouveaux protocoles d'évaluation biométrique, les bases d'apprentissage et les listes des tests sont fixées à l'avance.

méthode paramétrique L'hypothèse principale qui est faite lorsqu'on utilise cette méthode d'estimation de l'intervalle de confiance est que les scores sont issus de variables indépendantes et identiquement distribuées même pour les scores d'une personne, c'est-à-dire que tous les scores sont issus de la même loi et sont mutuellement indépendants. Cette hypothèse semble très stricte mais est nécessaire pour l'application des théorèmes. Ceci dit, elle n'est généralement pas vérifiée en pratique.

Il s'agit ensuite d'estimer les taux d'erreur $\text{FAR}(\theta)$ et $\text{FRR}(\theta)$ pour une ou plusieurs valeurs du seuil θ . Pour estimer ces erreurs, l'estimation des fonctions de répartition, appelées "fonctions de répartition empiriques", pour les deux classes "client" et "imposteur" doit être réalisée par une loi normale (voir [10]).

L'estimation des intervalles de confiance des taux d'erreur FAR et FRR à un seuil θ est obtenue par :

$$\begin{aligned}\text{FAR}(\theta) &\in [\widehat{\text{FAR}}(\theta) - \lambda * \sqrt{\frac{\widehat{\text{FAR}}(\theta)(1 - \widehat{\text{FAR}}(\theta))}{N_i}}, \widehat{\text{FAR}}(\theta) + \lambda * \sqrt{\frac{\widehat{\text{FAR}}(\theta)(1 - \widehat{\text{FAR}}(\theta))}{N_i}}] \\ \text{FRR}(\theta) &\in [\widehat{\text{FRR}}(\theta) - \lambda * \sqrt{\frac{\widehat{\text{FRR}}(\theta)(1 - \widehat{\text{FRR}}(\theta))}{N_c}}, \widehat{\text{FRR}}(\theta) + \lambda * \sqrt{\frac{\widehat{\text{FRR}}(\theta)(1 - \widehat{\text{FRR}}(\theta))}{N_c}}]\end{aligned}$$

avec $\widehat{\text{FAR}}(\theta)$ et $\widehat{\text{FRR}}(\theta)$ les valeurs calculées pour un test donné et N_c les scores "client" et N_i les scores "imposteur" du test. Pour des intervalles de confiance à 90%, on a $\lambda = 1,96$ et pour des intervalles de confiance de 95%, on a $\lambda = 1,645$.

Cette méthode non paramétrique sera utilisée dans la suite afin de calculer les intervalles de confiance. Ce choix est justifié par le nombre important de tests qui seront considérés pour la base FRGC, ce qui permet une bonne estimation des fonctions de répartition par une loi normale.

3.3.2 Évaluation des systèmes d'identification biométrique

Contrairement à la vérification, un seul type d'erreur peut survenir en mode identification à savoir le FMR (*False Match Rate*), c'est-à-dire dans le cas où il est assigné une fausse identité à la personne considérée, que cette personne fasse effectivement partie ou pas de la base de données. L'évaluation d'un système biométrique en mode identification est bien plus exigeante que celle des systèmes de vérification. Seuls des systèmes qui produisent des taux de EER faibles sont capables d'être déployés en mode identification. Ainsi, les protocoles d'évaluation diffèrent dans le mode identification et le mode vérification. Il faut savoir qu'en

mode identification, deux scénarios sont possibles qui induisent deux protocoles différents.

Si l'identification se fait dans ce qui est appelé un monde fermé, c'est-à-dire que chaque personne est représentée dans la base de données par au moins une référence, alors le système n'est pas obligé de faire du rejet. Il peut assigner l'identité de la personne qui a le score le plus élevé (dans le cas de mesure de score de similitude). L'évaluation doit se faire sur toutes les comparaisons possibles entre les éléments de la base des références et les autres éléments de la base afin de simuler au mieux le cas d'un processus d'identification.

Le système qui a le taux de FMR le plus faible est considéré comme le meilleur. L'évaluation peut se faire en utilisant des courbes de rang, c'est-à-dire que le FMR n'est pas mesuré sur la première décision mais sur k décisions avec k variant de 1 à N ; N étant inférieur ou égal au nombre de personnes enregistrées dans la base de référence. La figure 3.6 montre un exemple de courbe de rang tirée de l'évaluation FERET des systèmes de visage où le taux de bonne identification ($1 - \text{FMR}$) est représenté en fonction du rang pour trois algorithmes différents. Plusieurs interprétations de ces courbes sont possibles. En effet, nous pouvons considérer que le système, qui obtient le meilleur taux de FMR pour les rangs inférieurs ou égaux à N , est considéré comme le meilleur. Une autre stratégie consiste à comparer les systèmes à un rang précis k inférieur ou égal à N . En même temps, une troisième stratégie consisterait à considérer que le meilleur système est celui qui atteint sa meilleure performance le plus rapidement possible.

Un deuxième scénario est possible en mode identification, c'est le mode ouvert. Dans ce cas, toutes les personnes de la base de test ne sont pas représentées dans la base de référence. En mode évaluation, il est primordial de regrouper les personnes de la base de données considérée en deux groupes, celles disposant de références et celles qui n'en ont pas. Ensuite, le système doit pouvoir prendre deux décisions distinctes : soit rejeter un échantillon test ou soit lui assigner une identité représentée dans la base des référents. Pour mettre en place une stratégie de rejet, il est impératif de fixer un seuil de décision analogue à celui fixé en mode vérification.

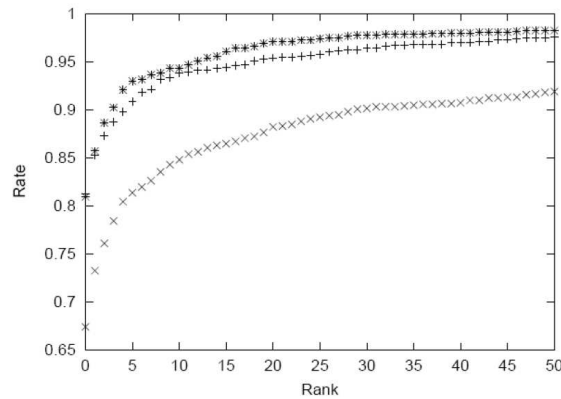


FIGURE 3.6: Courbe de rang en fonction du taux de bonne identification (Source NIST évaluation du visage FERET)

3.4 Bases de données publiques

3.4.1 Face Recognition Grand Challenge [NIST]

La base de données FRGCv2 est la plus importante base de données publique de visage [69]. Elle est composée de deux types d'images. Des images acquises dans des conditions d'illumination contrôlées et des images acquises dans des conditions d'illumination non contrôlées (respectivement FIG 3.7 (haut) et (bas)) et avec divers arrière plans (bureaux, couloirs,...). La totalité des images des visages sont prises de face.

La base FRGC est divisée en deux sous-ensembles :

- un ensemble d'images de développement formé de 12.776 images de 222 sujets, dont 6.388 images de visage prises dans des conditions non contrôlées
- un ensemble d'images de test formé d'images de 24.032 images de 466 sujets, dont 16028 images de visage prises dans des conditions contrôlées et 8014 images dans des conditions non contrôlées.

La figure 3.7 montre un exemple d'une session d'acquisition. Il est à noter que la base FRGCv2 contient des données 3D des visages, mais nous nous intéressons ici qu'à la partie des images 2D de cette base.

NB : Dans la suite de la thèse, nous utiliserons une sous-partie de la base de développement à des fins d'apprentissage. Cette sous-base qu'on nommera "BA" est constituée de 220 personnes avec 10 images de visage acquises dans des conditions contrôlées et 10 images acquises dans des conditions non contrôlées. Nous avons éliminé de la base deux personnes pour lesquelles le nombre d'images

acquises était insuffisant.

Protocoles d'évaluation Plusieurs expériences ont été définies dans le but d'étudier différents problèmes liés à la reconnaissance par le visage 2D et 3D [69]. Pour les images 2D, trois expériences sont définies².

- **Expérience 1** : Les ensembles de référence et de test sont composés d'images acquises dans des conditions contrôlées. Le modèle d'un sujet est constitué d'une seule image de référence et d'une seule image de test.
- **Expérience 2** : Les ensembles de référence et de test sont composés d'images acquises dans des conditions contrôlées. Le modèle d'un sujet est constitué de quatre images de référence et de quatre images de test.
- **Expérience 4** : L'ensemble des images de référence est composé d'images acquises dans des conditions contrôlées, alors que l'ensemble des images de test correspond à des images de visage acquises dans des conditions non contrôlées. Le modèle d'un sujet est constitué d'une seule image de référence et d'une seule image de test.



Images haute résolution type contrôlé



Images haute résolution type non contrôlé

FIGURE 3.7: Exemple de sessions d'acquisition de la base de données FRGCv2

2. L'expérience 3 de la base FRGC est une expérience qui concerne les données 3D

Enfin, afin de pouvoir analyser l'influence de la variation temporelle sur les performances des algorithmes pour chacune des trois expériences, des protocoles d'évaluation sont définis pour les trois expériences. Un premier protocole évalue les performances de comparaison d'images (références et tests) appartenant à des sessions d'acquisition d'un même semestre. Un deuxième protocole évalue les performances des tests d'image appartenant à des sessions d'acquisition de deux semestres consécutifs et un dernier, les performances des tests d'image de référence et de test, séparées d'une année. Les différents protocoles sont dénommés "masque" dans les évaluations sur la base FRGC.

Le tableau 3.1 résume les différents protocoles d'évaluation des trois expériences avec le nombre de tests "intra-classes" et "inter-classes" correspondants.

Expériences	Protocole	N°Tests intra-classes	N°Tests inter-classes
Expérience 1	1	196 672	55 217 232
	2	369 824	119 072 912
	3	173 152	63 855 680
Expérience 2	1	12 292	3 451 077
	2	23 114	7 442 057
	3	10 822	3 990 980
Expérience 4	1	98 336	27 608 616
	2	184 912	59 536 456
	3	86 576	31 927 840

TABLE 3.1: Nombre de tests pour les 3 protocoles dans les expériences 1, 2 et 4 de la base FRGC

Résultats de la littérature

Les premiers résultats sur la base FRGC ont été présentés lors du "Biometrics Consortium Conference" en 2005. Les résultats présentés par Jonathon Phillips montrent une grande différence entre les performances des deux types d'expériences (expérience 1 et expérience 4) (Figure 3.8). La première conclusion de cette présentation est que les conditions d'acquisition dégradées influencent de manière significative les performances. La deuxième conclusion est que les performances des systèmes de reconnaissance faciale ont bien progressé depuis la dernière campagne d'évaluations FRVT 2002.

Le meilleur résultat, présenté lors de ce consortium pour l'expérience 4, est donné par l'approche KFA (*Kernel Fisher Analysis*) appliquée aux représentations de Gabor du visage [49]. L'algorithme a atteint un taux de 78% de VR@0.1% de

FAR. Le deuxième meilleur résultat pour cette expérience est celui de la société Cognitec avec 62% de VR@0.1% de FAR. La médiane des résultats des algorithmes proposés est de 35% pour l'expérience 4, ce qui démontre la difficulté de cette expérience.

En ce qui concerne l'expérience 1, la plupart des algorithmes proposés ont des résultats corrects (lorsque les images de référence et de test sont de bonne qualité). Les résultats atteignent des taux de VR de 90% et 95% à 0.1% de FAR.

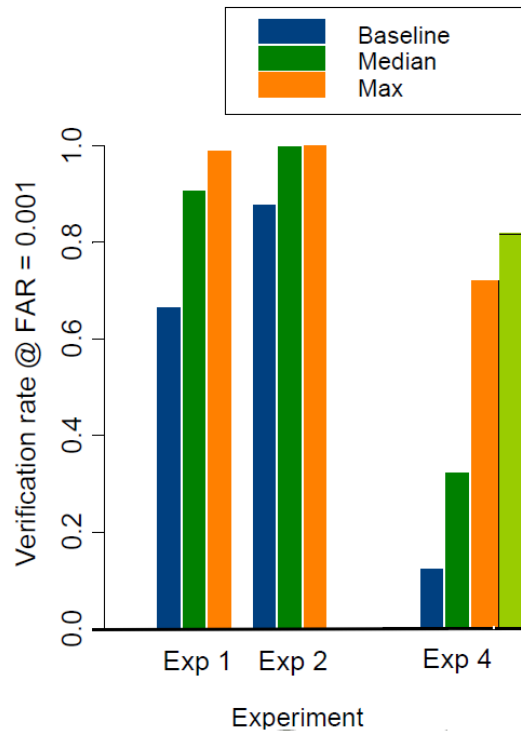


FIGURE 3.8: Premiers résultats obtenus lors des expériences de visage 2D de la base FRGCv2 (Source NIST Biometrics Consortium Conference)

Depuis 2005, des dizaines de travaux ont été publiés sur la base FRGCv2. Les meilleurs résultats de l'expérience 1 restent tous équivalents. Les résultats les plus intéressants restent ceux publiés sur l'expérience 4 par NIST.

En 2006, Savvides présente dans [79] ses résultats sur FRGC avec l'approche se basant sur les filtres de corrélation à noyaux. Les résultats sur l'expérience 4 ont atteint 72% de VR@0.1% de FAR. Cette méthode est une dérivée de l'approche MACE (*Minimum Average Correlation Energy*), présentée par le même auteur

dans [80]. Dans [29], l'auteur a rapporté un taux de VR@0.1% de FAR, qui est égal à 75.70%. La méthode utilisée est basée sur une réduction d'espace par LDA sur les réponses de trois filtres de Fourier à différentes bandes de fréquence.

En 2007, dans [88], l'auteur présente une approche basée sur une stratégie d'analyse et de fusion hiérarchique de plusieurs classifieurs. Ces classifieurs sont construits à partir des réponses des filtres de Gabor sur des parties de visage ainsi que sur la transformation de Fourier de l'image globale. Une réduction d'espace est appliquée en se basant sur une FDA (*Fisher Discriminant Analysis*). La classification finale est faite par une décision hiérarchique des différents classifieurs. L'auteur rapporte un taux de VR@0.1% de FAR de l'ordre de 86%.

3.4.2 BANCA [SURREY]

La base de données BANCA est une base qui contient deux modalités (visage et voix) [4] [55]. Deux types d'acquisition ont été effectués pour cette base : une acquisition des images et de la voix en haute définition et une acquisition en mode dégradé. Les personnes ont été enregistrées selon trois scénarios différents (contrôlés, dégradés et non contrôlés) comme présentés sur la figure 3.9, lors de 12 sessions différentes (4 sessions par scénario). Ces acquisitions s'étalent sur trois mois. Au total, les modalités de 208 personnes ont été capturées, la moitié sur des hommes et la moitié sur des femmes. Les données disponibles pour la communauté scientifique sont celles acquises sur 52 personnes, dont 26 sont des femmes et 26 sont des hommes.



FIGURE 3.9: Exemples d'image de la base de données BANCA selon trois scénarios : (a) contrôlé, (b) dégradé et (c) non contrôlé

Un nouveau point de fonctionnement a été introduit avec la base BANCA. C'est le point de fonctionnement WER (*Weighted Error Rate*), donné par la formule 3.5, correspondant à différentes conditions de fonctionnement. Ces conditions

correspondent à 3 différentes valeurs du rapport de coût de fausse acceptation (C_{FA}) et de faux rejet (C_{FR}), $r = C_{FA}/C_{FR}$, avec $r = 0.1$, $r = 1$ et $r = 10$.

$$WER(r) = \frac{P_{FA} + r * P_{FR}}{1 + r} \quad (3.5)$$

P_{FA} : probabilité de (FA) et P_{FR} : probabilité de (FR).

Protocoles d'évaluations

En utilisant les 12 sessions d'acquisitions, plusieurs protocoles ont été mis en place afin de comparer les différents algorithmes des participants à la compétition ICPR2004 [55].

Pour chacune des 52 personnes, 5 images sont utilisées pour les tests "clients" et 5 images sont utilisées pour les tests "imposteurs". Le modèle (référence ou test) de chaque sujet est donc constitué de 5 images et, à chaque sujet, il est assigné un "imposteur" particulier.

Deux groupes de test (G1 et G2) sont considérés. Chacun de ces groupes est constitué de 26 personnes (13 hommes et 13 femmes). Les paramètres des différents algorithmes sont alors optimisés sur le premier groupe G1 et sont ensuite utilisés pour le deuxième groupe G2 et vice-versa.

Plusieurs protocoles de tests sont proposés pour la base BANCA mais les deux protocoles P et Mc sont les plus fréquemment présentés dans la littérature. Le protocole P est le protocole le plus difficile du point de vue des conditions de test. En effet, les images de référence sont acquises dans des conditions contrôlées alors que celles des tests (client ou imposteur) sont acquises dans des conditions contrôlées, dégradées et diverses. Le protocole Mc est le protocole le plus facile car les images de références, ainsi que les images des tests "clients" et "imposteurs", sont acquises dans des conditions contrôlées.

Quelques résultats d'évaluation

Le tableau 3.4.2 résume les résultats du protocole P (extrait de [55]). L'algorithme de l'université de Tsinghua donne les meilleurs résultats, avec un taux moyen WER de 1.39%. L'approche utilisée se base sur une correction de l'illumination par lissage anisotropique et en appliquant une réduction d'espace par LDA sur l'amplitude des réponses de Gabor. Le système de référence de Biosecure [68], basé sur l'ACP sur les pixels de l'image, n'a pas participé à cette compétition

mais le résultat sur cette base a été rajouté dans le tableau pour avoir une idée sur la performance de l'ACP sur cette base.

	WER(0.1)		WER(1)		WER(10)		Av. WER %
	G1	G2	G1	G2	G1	G2	
IDIAP- HMM	8.69	8.15	25.43	20.25	8.84	6.24	12.93
IDIAP - FUSION	8.15	7.43	21.85	16.88	6.94	6.06	11.22
QUT	7.70	8.53	18.08	16.12	6.50	4.83	10.29
UPV	5.82	6.18	12.29	14.56	5.55	4.96	8.23
Univ Nottingham	1.55	1.77	6.67	7.11	1.32	1.58	3.33
Univ Taiwan	7.56	8.22	21.44	27.13	7.42	11.33	13.85
UniS	4.67	7.22	12.46	13.66	4.82	5.10	7.99
UCL - LDA	8.24	9.49	14.96	16.51	4.80	6.45	10.08
UCL - Fusion	6.05	6.01	12.61	13.84	4.72	4.10	7.89
NeuroInformatik	6.40	6.50	12.10	10.80	6.50	4.30	7.77
Tsinghua Univ	1.13	0.73	2.61	1.85	1.17	0.84	1.39
CMU	5.79	4.75	12.44	11.61	6.61	7.45	8.11
BIOSECURE RefSys	8,95	10,23	26,85	26,59	8,35	6,62	14,60

TABLE 3.2: Résultats du protocole P sur la base BANCA extrait de [55]

NB : Dans la suite de cette thèse, nous utiliserons cette base afin de valider nos résultats obtenus. Les protocoles P et Mc seront les plus souvent invoqués.

3.4.3 IV²

Le projet IV² est un projet d'évaluation biométrique français. Il a pour but de créer des ressources et des conditions d'une évaluation à l'échelle nationale et internationale de différents systèmes liés à l'information du visage, de l'iris et de la voix, dans des milieux semi-contraints.

Pour constituer la base de données IV², une cabine d'acquisition très semblable à une cabine de type photomaton a été utilisée et est représentée dans la figure 3.10.

Afin d'obtenir un éclairage optimal dans la cabine, nous disposons de plusieurs lampes halogènes qui seront pilotées depuis un ordinateur à l'aide de télé-variateurs de puissance.

La base IV² est une base multimodale comportant une séquence de 2 minutes de visage parlant qui est acquise avec des caméras différentes dont une "Webcam", des données du visage 3D avec différentes expressions et sous diverses conditions d'illumination, des données de visage acquises par stéréovision et des images d'iris

acquises en proche infrarouge.

Cette base de données a été acquise sur trois sites différents : l'INT (EVRY), l'université d'Evry et l'école Centrale de Lyon (LYON).

Protocoles d'évaluation

La base d'évaluation comporte 482 identifiants, correspondant à 315 personnes dont 268 personnes sont enregistrées sur une session, 77 enregistrées sur 2 sessions, et 19 enregistrées sur 3 sessions. Sa taille est de 19 Go, correspondant à 7651 dossiers et 56111 fichiers.

Les tests se font sur des images fixes, extraites des vidéos. L'extraction se fait de manière automatique. Les images sont au format jpg, avec un facteur de qualité de 95. Les collections d'image sont réalisées de la manière suivante :

- images camscope ("dv") :
 - première séquence : 8 images, espacées de 4 secondes, à partir de l'instant 0 seconde
 - deuxième séquence : 8 images, espacées de 4 secondes, à partir de l'instant (fin de l'enregistrement - 37 secondes)
- images caméra numérique ("dcam") :
 - première séquence : 8 images, espacées de 1 seconde, à partir de l'instant 2 secondes
 - deuxième séquence : 8 images, espacées de 2 secondes, à partir de l'instant 26 secondes
- images webcam ("wcam") : une séquence de 8 images, espacées de 5 secondes, à partir de l'instant 1 seconde

Les données disponibles sont utilisées pour construire plusieurs séries de tests, présentées dans le tableau 3.3 :

Experience	2D-Exp1	2D-Exp2	2D-Exp3	2D-Exp4
Sessions	Mono	Mono	Mono	Multi
qualité	Haute	Haute	Basse	Haute
Expression V.	faible	Non	Non	Non
Illumination V.	Non	Oui	Non	Non
N. Intra-classe	2595	2502	1654	1796
N. Interclasse	2454	2362	1598	1796

TABLE 3.3: Protocoles d'évaluation pour la base IV²

Trois algorithmes ont été comparés dans cette évaluation, dont celui que j'ai

Vue externe de la cabine IV²Vue interne de la cabine IV²

Exemple d'image d'iris



Exemple d'images par stéréovision



Exemple d'acquisition 3D

FIGURE 3.10: Exemples de données de la base IV²



FIGURE 3.11: Exemples d'acquisition d'images 2D : (a) image haute résolution et (b) image dégradée

développé dans cette thèse, c'est pour cette raison que j'ai choisi de présenter les résultats de cette évaluation dans le chapitre 6 où je décris mon approche.

3.5 Conclusion

Dans ce chapitre, nous avons listé les outils nécessaires à l'évaluation des performances des algorithmes biométriques en identification et en vérification ainsi que les bases de données publiques de visage (FRGC, BANCA et IV²) et les protocoles de tests associés. Ces bases et leurs protocoles seront utilisés tout au long du reste de cette thèse.

Chapitre 4

Les algorithmes de reconnaissance de visage

4.1 Introduction

Plusieurs approches pour la reconnaissance par le visage ont été présentées dans le chapitre 2. Comme nous l'avons précisé dans ce chapitre, il y a deux grandes familles : les algorithmes basés sur les approches globales par réduction d'espace et les algorithmes basés sur les approches par points caractéristiques. Nous allons nous limiter à la première famille étant donné que nous allons utiliser dans notre travail une méthode globale.

Dans ce chapitre, nous présenterons donc trois algorithmes de réduction d'espace : l'ACP, la LDA et la DLDA. Une présentation théorique de ces approches, et surtout leur application au domaine du visage, est faite dans la première partie de ce chapitre. Une étude comparative entre ces algorithmes est présentée dans la deuxième partie de ce chapitre.

Nous nous focaliserons ensuite sur plusieurs particularités que rencontrent ce type d'algorithmes comme le choix des conditions adéquates d'apprentissage. L'objectif de ce chapitre est de comparer les différentes méthodes afin de pouvoir choisir le meilleur algorithme pour la suite de nos études.

4.2 Dimensionnalité de l'espace visage

Une image du visage est un signal à 2 dimensions, acquis par un capteur digital (caméra numérique, scanner...). Ce capteur codera la couleur ou l'intensité des différents points de l'image dans une matrice de pixels à deux dimensions¹.

Après une normalisation de l'image et une mise à l'échelle vers une taille fixe (ex : $m * n$), l'image du visage peut être considérée comme un vecteur dans un espace multi-dimensionnel \mathbb{R}^{mn} . Ce vecteur est obtenu en mettant, simplement, dans un ordre fixe les éléments de la matrice des pixels. Afin de définir une image (ou un point) dans l'espace des images, nous devons spécifier une valeur pour chaque pixel de cette image. Le nombre de points constituant cet espace devient rapidement très grand, même pour les images de petite dimension. Cette dimensionnalité pose un certain nombre de problèmes pour les algorithmes de reconnaissance, qui se basent sur cette représentation de l'image, à savoir :

- dans un contexte de la reconnaissance, travailler dans un grand espace pose un problème de complexité de calcul.
- pour les méthodes paramétriques, le nombre de paramètres à estimer peut rapidement dépasser le nombre d'échantillons d'apprentissage, ce qui pénalise l'estimation.
- pour les méthodes non paramétriques, le nombre d'exemples nécessaires afin de représenter efficacement la distribution des données peut être insuffisant.

Le cas du visage est assez particulier. Le visage est formé par des surfaces lisses et une texture régulière. Il possède une forte symétrie (dans le cas des images frontales) et il est formé à partir de mêmes objets (yeux, nez et bouche), qui ont les mêmes positions pour tous les visages. Ces spécificités donnent lieu à plusieurs constatations :

- les différents pixels de l'image du visage sont fortement corrélés.
- les images des différents visages sont aussi corrélées.
- un important nombre de points dans l'espace des images ne représentent pas des visages.

D'après ces constatations, nous pouvons déduire que les visages forment une classe particulière dans l'espace des images. L'ensemble des images de visage forment alors un sous-espace de l'espace des images. D'une façon plus formelle, si on considère Γ l'espace des images 2D d'une taille fixe, $l = n * m$. On a $\Gamma \subset \mathbb{R}^l$. No-

1. Dans ce chapitre, nous allons nous intéresser à l'image de l'intensité de l'illumination, appelée aussi image en niveaux de gris

tons F l'espace des images des visages normalisés à une taille $n * m$. Nous avons $F \subset \mathbb{R}^l, F \subset \mathbb{R}^l$.

En 1994, Ruderman a démontré que les images naturelles possèdent une grande redondance statistique [77]. En 1996, Penev [65] a démontré que dans le cas précis des images normalisées des visages, cette redondance statistique est d'autant plus forte.

L'appartenance à un même sous-espace et la redondance statistique qui caractérise les visages expliquent l'intérêt des chercheurs pour les méthodes de reconnaissance basées sur les approches par réduction d'espace, appelées également "les approches globales". Le fait que les visages soient confinés dans un sous-espace permet de définir une base (orthogonale, ou non orthogonale) dans cet espace. Cette base devrait représenter de façon plus pertinente n'importe quel point de l'espace des visages. L'avantage de la redondance statistique est qu'elle permet une extraction d'une structure simple des caractéristiques importantes et pertinentes de l'image du visage. Cette structure permettrait de représenter le visage tout en gardant l'information la plus importante, et par conséquent, de réduire la dimensionalité de l'espace visage. Tout l'intérêt des approches globales est la construction de cette base de projection qui permettra de comparer, de reconnaître ou d'analyser l'information essentielle des visages.

4.3 Analyse en composantes principales

En 1987 et 1990, Sirovich et Kirby [39, 87] publient leurs travaux sur la reconstruction de visage à partir de vecteurs de projection par ACP (Analyse en Composantes Principales) ou KL (*Karhunen-Loève*). Ils montrent la validité de la réduction de dimensionalité de l'espace des visages et aussi qu'une reconstruction de visage à partir de la nouvelle base est même meilleure sur le plan visuel, car elle corrige une grande partie des bruits ou des zones occultés lors de la capture. Cette approche a été reprise par Turk en 1991 [93] et appliquée à la reconnaissance des visages. Cette méthode reste d'actualité et même considérée comme une méthode de base. Elle est intégrée dans les deux systèmes de référence de FRGCv2 [69] et de Biosecure [53, 68].

4.3.1 Réduction d'espace par ACP

L'idée fondamentale de l'ACP est la suivante : considérant un nuage de N points en P dimensions, on cherche à trouver le plan dans lequel la projection des points du nuage est la moins déformée possible, donc la plus fiable possible.

Quand on projette un ensemble de points dans un espace ayant moins de dimensions (le cas pour toutes les méthodes de réduction d'espace), la distance entre les points ne peut que diminuer. Pour trouver le meilleur plan de projection, on cherche le plan dans lequel la distance entre les points projetés demeurent en moyenne maximale. Ce plan est qualifié de "principal".

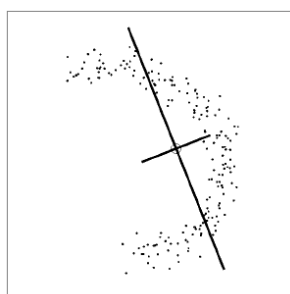


FIGURE 4.1: Exemple de distribution d'un nuage de points dans un espace 2D

Plus généralement, on définit P droites orthogonales les unes aux autres qui permettent de définir un repère orthonormé. Ces P droites sont les P "axes principaux" d'un repère dans lequel sont situés les échantillons de manière à les décrire de la façon la plus concise. L'objectif est que la plupart des coordonnées d'un échantillon soient presque nulles et que seules quelques coordonnées aient une valeur importante.

Les coordonnées d'un échantillon i de coordonnées $(a_{j,j \in 1, \dots, p})$ dans ce nouveau repère s'expriment par rapport à de nouveaux attributs appelés "composantes principales". Notons $z_{j \in 1, \dots, p}$ les composantes principales, $z_{i,j}$ dénotant le j -ème caractère principal de l'échantillon i . On a naturellement une relation reliant les attributs originaux aux nouveaux attributs.

La meilleure représentation des données au moyen de $q < P$ attributs s'obtient en ne prenant en compte que les q premières composantes principales. L'ACP est une méthode factorielle car elle construit de nouvelles coordonnées par combinaison des coordonnées initiales.

Les N échantillons décrits par leur P attributs peuvent être mis sous forme

d'une matrice ayant N colonnes et P lignes : 1 colonne décrit un échantillon et chaque ligne correspond à un attribut. Notons cette matrice \mathbf{X} . On définit la matrice de variance \mathbf{V} de la manière suivante :

$$\mathbf{V} = \begin{pmatrix} \text{var}(a_1) & \text{covar}(a_1, a_2) & \dots & \text{covar}(a_1, a_p) \\ \text{covar}(a_2, a_1) & \text{var}(a_2) & \dots & \text{covar}(a_2, a_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{covar}(a_p, a_1) & \text{covar}(a_p, a_2) & \dots & \text{var}(a_p) \end{pmatrix} \quad (4.1)$$

et la matrice de corrélation \mathbf{R} :

$$\mathbf{R} = \begin{pmatrix} 1 & r(a_1, a_2) & \dots & r(a_1, a_p) \\ & 1 & \dots & r(a_2, a_p) \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \quad (4.2)$$

qui est symétrique puisque $r(a_j, a_k) = r(a_k, a_j)$.

Pour des échantillons centrés et réduits, ces deux matrices sont égales $\mathbf{R} = \mathbf{V}$. On a la relation : $\mathbf{R} = \frac{1}{P} \mathbf{X} * \mathbf{X}^T$. L'analyse de la matrice de corrélation est intéressante car elle permet de repérer immédiatement les caractères fortement corrélés et ceux qui ne le sont pas.

Le calcul des valeurs propres et des vecteurs propres de \mathbf{R} fournit alors toutes les informations recherchées. Notons $\lambda_{i \in 1, \dots, p}$ les P valeurs propres de \mathbf{R} ordonnées de manière décroissante $\lambda_1 > \lambda_2 > \dots > \lambda_p$, et $\vec{V}_{i \in 1, \dots, p}$ leurs P vecteurs propres associés. \mathbf{R} est par nature symétrique et définit positive. Ses valeurs propres sont réelles et positives et ses vecteurs propres ont des coordonnées réelles. S'il existe une ou des valeurs propres nulles, cela signifie que les attributs ne sont pas linéairement indépendants les uns des autres : un ou plusieurs attributs sont obtenus par combinaison linéaire des autres.

Ces vecteurs propres sont unitaires et orthogonaux deux à deux. Ils forment donc une base orthonormée. Ce sont les axes principaux recherchés. De plus, l'axe principal est celui associé à la valeur propre la plus grande. Ainsi, les q axes principaux sont les droites dont les vecteurs unitaires sont les vecteurs propres associés aux q valeurs propres les plus grandes. On obtient alors les coordonnées principales des échantillons en les projetant dans cet espace, c'est-à-dire en faisant le produit scalaire des coordonnées (centrées réduites) d'un échantillon par chacun des q vecteurs propres choisis.

Sélection des vecteurs propres

L'ordre de grandeur des valeurs propres les unes par rapport aux autres indique leur importance. Si quelques valeurs propres ont des valeurs bien plus importantes que les autres, cela signifie que l'essentiel des informations est donnée par ces axes principaux et que les autres axes donnent peu d'information. Au contraire, si toutes les valeurs propres ont sensiblement la même valeur, alors aucun axe principal ne peut être laissé de côté. Voulant extraire l'information essentielle, la situation idéale est lorsque q valeurs propres sont très importantes par rapport aux autres. Il est facile de démontrer que $\sum_{i \in 1, \dots, P} \lambda_i = P$. L'importance d'une valeur propre par rapport aux autres est alors mesurée par son inertie :

$$I_i = \frac{\lambda_i}{P} \quad (4.3)$$

où λ_i est la variance le long de l'axe principal porté par le vecteur propre \vec{V}_i donc c'est la variance de la composante principale z_i . On définit aussi la variance de l'espace réduit à la dimension q par $\sum_{i \in 1, \dots, q} \lambda_i$.

Turk et Pentland ont appliqué l'ACP (Analyse en Composantes Principales) au problème de reconnaissance de visage en 1991 [93], connue aussi sous le nom de méthode de Kerhunen-Loève.

Afin de pouvoir appliquer l'ACP à la modalité visage, les images de la base d'apprentissage de l'espace de réduction doivent être normalisées géométriquement. En d'autres termes, nous transformons la région du visage afin d'avoir un même nombre de pixels pour tous les visages de la base. Dans la littérature, une distance fixe entre les centres des yeux pour l'image de sortie est imposée [11].



FIGURE 4.2: Exemple d'image de visage normalisé (nombre de pixels fixe)

On cherche les axes qui maximisent la variance entre ces visages. Les n images sont donc transformées d'une matrice 2D en un vecteur de dimension l afin de construire la matrice de données $M \in \mathbb{R}^{ln}$. On calcule la matrice de covariance C_M

ainsi que la moyenne μ_M . Toutes les images de la base sont centrées par rapport à μ_M . Afin de retrouver les axes de la base orthogonale (base de projection), le problème est ramené au calcul du problème des valeurs propres :

$$C_M V_i = \lambda_i * V_i \quad (4.4)$$

avec λ_i valeur propre correspondant au vecteur propre V_i . L'ensemble des vecteurs propres, correspondants aux valeurs propres non nulles, formeront une base orthogonale.

Les valeurs propres sont ensuite classées dans l'ordre décroissant et seules sont gardées les vecteurs propres correspondants à une certaine valeur de la variance. Dans la pratique, on garde les vecteurs propres qui conservent les 95% de la variance de l'espace. On notera par la suite $\psi = \{V_i\}_{i \in \{1, \dots, r\}}$, avec ψ la matrice de transformation de l'espace d'entrée de dimension l vers l'espace réduit de dimension r avec $l \gg r$.

4.3.2 Application de l'ACP à la reconnaissance de visage

Si on considère deux images des visages (ϕ^1, ϕ^2) normalisées et centrées (moyenne nulle et variance égale à 1 pour les valeurs des pixels), la comparaison entre les deux images est faite par une mesure métrique entre les projetés des images dans l'espace réduit : $d(\phi^1, \phi^2) = \|\psi(\phi^1), \psi(\phi^2)\|$, avec $\|\cdot\|$ la mesure métrique considérée. Différentes mesures ont été proposées dans la littérature, comme la distance L1, la distance euclidienne L2, la similarité cosinus (ou "mesure cosinus"), la distance de Mahalanobis. Voici leurs définitions :

– **L1** :

$$d(\phi^1, \phi^2) = |\psi(\phi^1) - \psi(\phi^2)| = \sum_{i=1}^r |a_i - b_i| \quad (4.5)$$

avec a_i et b_i les coordonnées de projection de $\psi(\phi^1)$ et $\psi(\phi^2)$

– **L2** :

$$d(\phi^1, \phi^2) = \|\psi(\phi^1) - \psi(\phi^2)\| = \sum_{i=1}^r (a_i - b_i)^2 \quad (4.6)$$

– **Angle** :

$$d(\phi^1, \phi^2) = -\frac{\psi(\phi^1) - \psi(\phi^2)}{\|\psi(\phi^1)\| \|\psi(\phi^2)\|} = -\frac{\sum_{i=1}^r a_i b_i}{\sqrt{\sum_{i=1}^r (a_i)^2 \sum_{i=1}^r (b_i)^2}} \quad (4.7)$$

– **Mahalanobis** :

$$d(\phi^1, \phi^2) = -\sum_{i=1}^r \frac{1}{\sqrt{\lambda_i}} a_i b_i \quad (4.8)$$

Moon et Phillips ont étudié dans [58] les performances de ces 4 mesures sur la base FERET [32]. Une autre étude [95], faite en 2002, a repris les travaux de Moon et Phillips et a confirmé aussi que la distance de Mahalanobis dépassait les autres distances en terme de performance et qu'une fusion de ces mesures améliorerait la reconnaissance. Ces évaluations ont été faites sur la même base FERET.

Dimensionnalité optimale de l'espace de réduction construit par ACP

Différentes études de la littérature [32, 64, 95] proposent de fixer de manière empirique la variance de l'espace de réduction entre 92% et 98%. Cet intervalle est bien justifié dans la reconstruction des images par ACP mais ne l'est pas dans le contexte de reconnaissance. Le but de l'utilisation de l'ACP en reconstruction d'image est d'avoir une erreur quadratique moyenne minimale entre l'image d'origine et l'image reconstruite. En reconnaissance, la réduction d'espace n'est pas liée à cette même erreur quadratique mais à la quantité d'information utile que la nouvelle représentation peut avoir.

Afin de déterminer la dimensionnalité de l'espace de réduction construit par ACP, Penev et Sirovich dans [64] ont utilisé des critères visuels de reconstruction. Ils ont affirmé que pour des images de visage de taille réduite (~ 25) de pixels entre les yeux, la dimensionnalité adéquate pour une bonne reconnaissance se situe entre 400 et 700 vecteurs propres.

Dans [54], nous avons proposé une méthode de sélection de la dimension optimale de l'espace de réduction en se basant sur l'étude du rapport entre l'énergie de l'information et l'énergie de l'information résiduelle. Cette approche est intéressante dans le sens où elle montre une corrélation entre ce rapport d'énergies et les performances du système de reconnaissance.

Soit PW l'énergie de l'espace construit avec toutes les composantes principales de l'espace :

$$PW = \sum_{i=1}^n \lambda_i^2 \quad (4.9)$$

avec n le nombre total des valeurs propres de λ .

Soit S_r l'apport énergétique de la composante principale r définit par :

$$S_r = \frac{\lambda_r^2}{PW} \quad (4.10)$$

Le rapport R_r entre l'énergie résiduelle à partir de la composante principale r et l'énergie totale du système PW est défini par :

$$R_r = \frac{\sum_{i=r}^n \lambda_i^2}{PW} \quad (4.11)$$

On considère le rapport $\mathbf{v}(r)$ entre l'apport énergétique de la composante principale r et l'énergie résiduelle à partir de cette composante. On appellera $\mathbf{v}(r)$ le taux d'énergie résiduelle exprimé par l'équation suivante :

$$\mathbf{v}(r) = S_r/R_r = \frac{\lambda_r^2}{\sum_{i=r}^n \lambda_i^2} \quad (4.12)$$

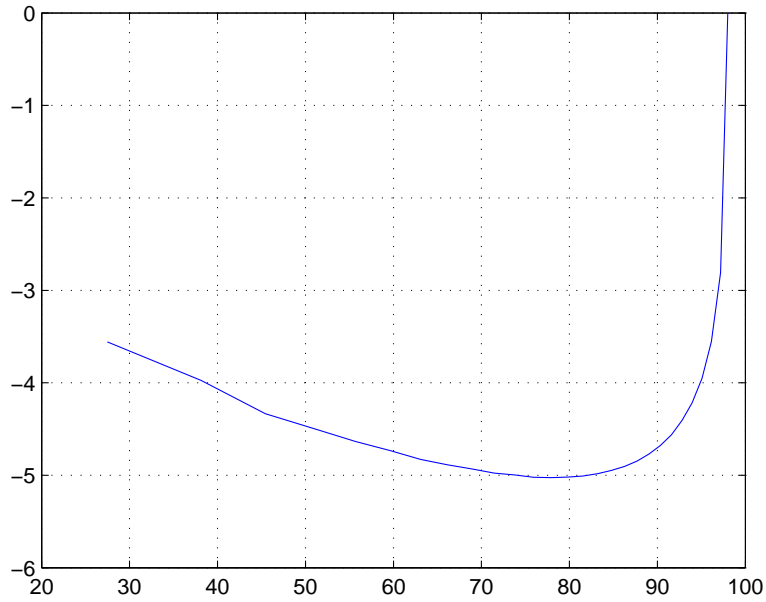


FIGURE 4.3: Évolution de la mesure du taux d'énergie résiduelle en fonction de la variance de l'espace pour la sous-base apprentissage de FRGC

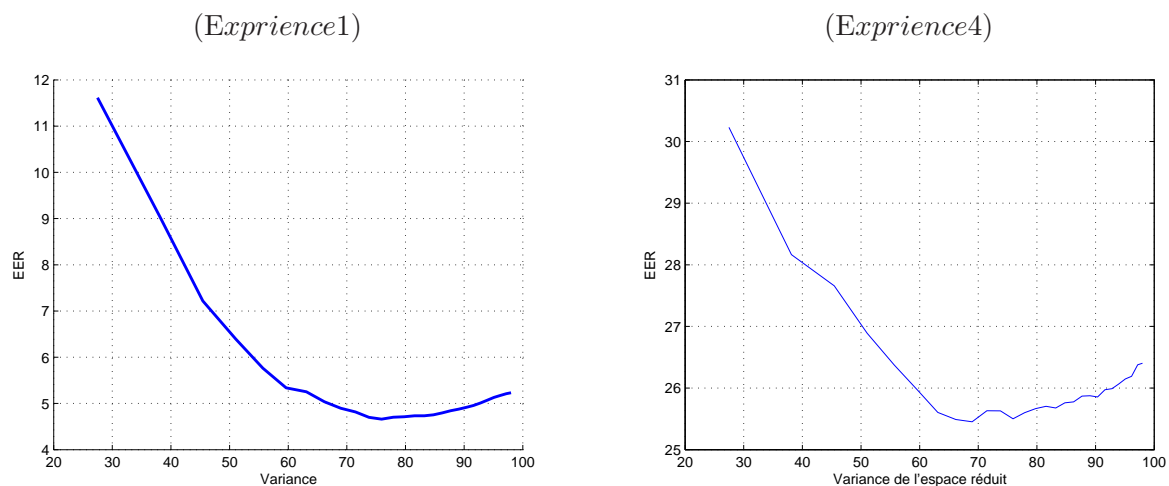


FIGURE 4.4: Évolution du EER en fonction de la variance pour l'expérience 1 et l'expérience 4 de FRGC

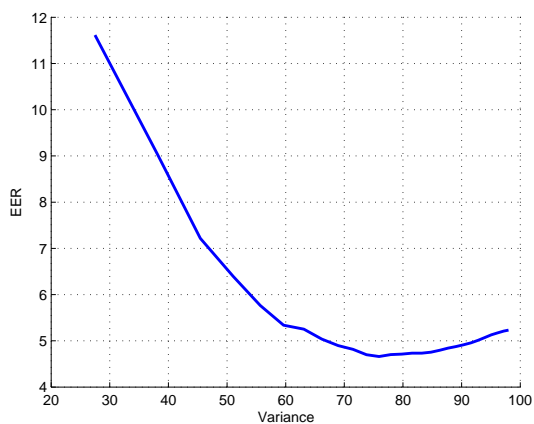


FIGURE 4.5: Évolution du EER en fonction de la variance pour le protocole P de BANCA

Les résultats des évaluations sur la base FRGC (figure 4.4) et ceux sur la base BANCA (figure 4.5) montrent que l'utilisation des premiers axes principaux seuls détériore les résultats de la reconnaissance (les axes principaux qui expriment moins de 40% de la variance de l'espace) . Cette détérioration est due au fait que l'information discriminante se situe non seulement dans les premiers axes principaux mais aussi dans le reste des axes. En plus, l'ajout d'un surplus d'axes principaux peut aussi dégrader la reconnaissance (Figure 4.3 (Les axes à faible apport de variance qui donnent une variance supérieure à 75%)).

Dans cette étude, nous nous sommes aussi intéressés à l'influence des échantillons d'apprentissage sur les performances de la reconnaissance de visage. Nos expériences ont montré également que la présence des variabilités comme l'expression ou l'illumination dans la base d'apprentissage de l'espace de réduction est plus importante que la variabilité des personnes.

On considère la base de développement² de FRGC et un sous-ensemble de la base BANCA. Les sous-ensembles d'apprentissage construits à partir de ces deux bases sont résumés ci-dessous et dans le tableau 4.1 :

- F1 : un sous-ensemble de la base d'apprentissage (BA) de FRGC, construit avec 18 personnes et 512 images acquises en conditions contrôlées et 512 en conditions non contrôlées
- F2 : un sous-ensemble de BA de FRGC, construit avec 35 personnes et 1024 images en conditions contrôlées et 1024 en conditions non contrôlées.
- F3 : un sous-ensemble de BA de FRGC, construit avec 222 personnes et 1110 images de conditions contrôlées et 1110 en non contrôlées.
- B1 : l'ensemble de la base BANCA , construit avec 52 personnes et 1020 images dont 40% sont acquises en conditions contrôlées, 40% en non contrôlées et 20% en dégradées.

Le sous-ensemble d'apprentissage de la base de BANCA a été choisi afin de mesurer les performances de l'ACP lorsque l'espace de réduction est construit à partir d'une base d'apprentissage complètement différente de la base d'évaluation (en termes de personnes). On peut aussi voir l'impact de l'utilisation de plusieurs bases d'apprentissage sur les performances de la reconnaissance (F1 + B1, union des deux bases F1 et B1).

La dimensionalité de l'espace de réduction est choisie par rapport à la zone de

2. Dans le chapitre précédent, nous avons présenté la base FRGC 3.4. La base d'apprentissage (BA) est un ensemble d'images, formée de 12.776 images de 222 sujets dont 6.388 images de visage prises dans des conditions non contrôlées.

BA	Sous-ensemble	N°d'images	N°de personnes	C/NC/D
FRGCv2	F1	1024	18	50% / 50% / 0%
-	F2	2048	35	50% / 50% / 0%
-	F3	2220	222	50% / 50% / 0%
BANCA	B1	1020	52	40% / 40% / 20%

C : Images acquises en conditions contrôlées

NC : Images acquises en conditions non contrôlées

D : Images acquises en conditions dégradées

TABLE 4.1: Caractéristiques des sous-ensembles d'apprentissage de l'espace de réduction par ACP

stabilité du taux d'énergie résiduelle qui se situe à $\sim 80\%$ de la variance de l'espace total. Le tableau 4.2 donne les résultats de l'expérience 4 de FRGC suivant les différents espaces de réduction construits à partir des bases d'apprentissage du tableau 4.1.

	B1	F1	F1 + B1	F2	F3
EER	28,2%	23,24%	22,9%	23,4%	26,9%
IC à 90%	[$\pm 2.1\%$]	[$\pm 1.4\%$]	[$\pm 1.4\%$]	[$\pm 1.4\%$]	[$\pm 1.9\%$]

TABLE 4.2: Résultats en EER de l'expérience 4 de FRGC pour les différentes bases d'apprentissage (IC : Intervalle de Confiance).

Le premier point notable de ces expériences est que l'apprentissage de l'ACP pose un problème de généralisation, retrouvé dans toutes les méthodes de reconnaissance par réduction d'espace. Sachant que si l'espace de réduction est construit à partir de personnes qui ne sont pas dans la base d'évaluation FRGC, les performances sont dégradées de façon significative. Pour ces expériences le taux d'erreur EER passe de 28.2% à 23.4%.

On remarque que la variabilité des personnes est moins importante que la variabilité des situations environnementales. En effet, pour F3, nous avons 220 personnes avec 5 images de type non contrôlé par personne alors que pour F1, nous avons 18 personnes avec en moyenne 28 images de type non contrôlé. Les performances de la PCA passent de 26% en utilisant F3 à 23.2% en utilisant F1.

4.4 Fisher Discriminant Analysis

Dans le cas où on possède l'information d'appartenance des données à des classes prédéfinies, il est intéressant d'utiliser une approche qui permet de prendre

en compte cet élément, ce que l'ACP ne permet pas de faire . En 1936, Fisher propose une méthode de classification basée sur la maximisation de la distance "inter-classes" et la minimisation de la distance "intra-classes".

4.4.1 Analyse Linéaire Discriminante

L'Analyse Linéaire Discriminante (LDA, *Linear Discriminant Analysis* ou *Fisher Discriminant Analysis*) permet de rechercher, dans l'espace des données, les axes qui permettent de discriminer au mieux les différentes classes (plutôt que les axes qui décrivent au mieux les données, tel est le cas de l'ACP). En d'autres termes, si on considère un certain nombre de paramètres indépendants, la LDA permet de créer une combinaison linéaire de ces derniers qui aboutit à un maximum de distance entre les moyennes des différentes classes. Concrètement, pour tous les échantillons de toutes les classes, on définit deux mesures. La première mesure est la matrice d'éparpillement intra-classes S_w ("*within-class scatter matrix*") qui est définie par :

$$S_w = \sum_{j=1}^c \sum_{i=1}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T \quad (4.13)$$

avec x_i^j le i ème échantillon de la classe j , μ_j la moyenne de la classe j , c le nombre de classes et N_j le nombre d'échantillons de la classe j . La deuxième mesure est la matrice d'éparpillement inter-classes S_b ("*between-class scatter matrix*") qui est définie par :

$$S_b = \sum_{j=1}^c (\mu_j - \mu)(\mu_j - \mu)^T \quad (4.14)$$

avec μ la moyenne de tous les échantillon.

Le but est de maximiser les distances inter-classes tout en minimisant les distances intra-classes, ce qui revient à retrouver la matrice de transformation W qui maximise le critère $J(W) = \frac{W^T S_b W}{W^T S_w W}$ [21] donc W est optimale pour :

$$W_{opt} = \arg \max_W \left(\frac{|W^T S_b W|}{|W^T S_w W|} \right) = [w_1, w_2, \dots, w_m]$$

Ce problème est ramené à un problème de recherche des vecteurs propres de la matrice $S_w^{-1} S_b$.

La figure 4.6 représente un exemple de classification de deux nuages de points. L'axe principal de la LDA est l'axe de projection qui maximise la séparation entre

les deux classes. Il est clair que cette projection est optimale pour la séparation des deux classes par rapport à la projection sur l'axe principal calculé par ACP.

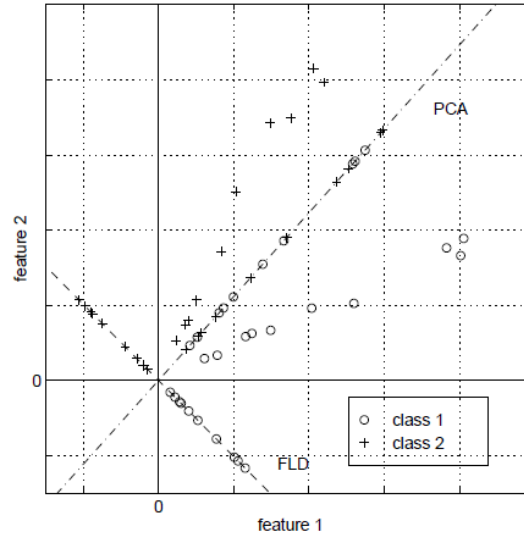


FIGURE 4.6: Comparaison entre les projections de deux classes de points ("class 1" et "class 2") sur les axes principaux construits par ACP et par FLD=LDA (Figure tirée de [7])

La maximisation de $J(W)$ n'est possible que si la matrice S_w est singulière (inversible). Cette condition pose un problème pour les applications de reconnaissance du visage³.

Pour remédier à ce problème, Swets [89] a proposé d'utiliser l'espace ACP réduit comme espace intermédiaire. On commence par rechercher les axes principaux des données par la ACP. On réduit les données de départ dans cet espace (projection dans une base orthogonale) et enfin, on applique la LDA sur ces données réduites.

4.4.2 La LDA directe (DLDA)

L'idée de la LDA est de maximiser le critère de Fisher. Cela revient à rechercher une matrice A qui diagonalise simultanément les deux matrices d'éparpille-

3. Le nombre d'images est généralement très inférieur à la dimension de l'image donc la matrice S_w est de rang inférieur à C donc non inversible. Ce problème est connu sous le nom de problème de la taille réduite des échantillons *SSS* ("Small Sample Size".)

ment S_w et S_b :

$$AS_wA^T = I, \quad AS_bA^T = \Lambda \quad (4.15)$$

Avec Λ matrice diagonale ordonnée par ordre décroissant.

Dans [16], Chen a démontré que l'espace nul de S_w ⁴ contient de l'information discriminante et qu'un passage par une réduction d'espace par ACP supprime cette information. Concrètement, si pour un vecteur " a " de l'espace d'entrée, $S_w a = 0$, et $S_b a \neq 0$, alors $\frac{|a^T S_b a^T|}{|a^T S_w a^T|}$ est maximisée. Il suffit alors de garder l'espace nul de la matrice d'éparpillement inter-classes S_w . Cette méthode est bien meilleure que la LDA classique dans le sens où les vecteurs de projection obtenue maximisent le critère de Fisher sans supprimer l'information discriminante.

Une nouvelle approche, appelée la DLDA directe, a été proposée dans [103] afin de maximiser le critère de Fisher. La première étape de cette approche est la diagonalisation de la matrice S_b par le calcul de la matrice V :

$$V^T S_b V = \Lambda$$

avec $V^T V = I$. Le problème est alors ramené à la résolution d'un problème de valeurs propres. Chaque vecteur de la matrice V est un vecteur propre de la matrice S_b et Λ contient toutes les valeurs propres de S_b . Comme la matrice S_b peut être singulière, Λ peut contenir des valeurs propres nulles ou très faibles. Ces valeurs ainsi que les vecteurs propres associés doivent être écartés car une projection dans la direction de ces vecteurs n'apporte aucune information discriminante entre les classes.

Soit Y la sous-matrice composée des m premières colonnes de V (Y étant de dimension $m \times n$ et n étant la taille de l'échantillon d'entrée) :

$$Y^T S_b Y = D_b > 0$$

avec D_b la sous-matrice de Λ , de dimension $m \times m$.

En multipliant par $D_b^{-1/2T}$ et $D_b^{-1/2}$ de part et d'autre on aboutit à :

$$(YD_b^{-1/2})^T S_b (YD_b^{-1/2}) = I$$

En posant $Z = YD_b^{-1/2}$, nous obtenons :

$$Z^T S_b Z = D_b \quad (4.16)$$

4. L'espace nul de $S_w = \{x | S_w x = 0, x \in \mathbb{R}^n\}$.

On constate que Z permet de réduire la dimension de S_b de $n \times m$ à $m \times m$.

Considérons la diagonalisation de la matrice $Z^T S_w Z$ par la résolution du problème des valeurs propres :

$$U^T Z^T S_w Z U = D_w \quad (4.17)$$

avec $U^T U = I$. D_w peut contenir des valeurs nulles sur la diagonale.

L'objectif est de minimiser l'éparpillement des inter-classes. Il est donc important de garder les vecteurs de projection associés aux valeurs propres les plus faibles, spécialement les valeurs nulles, et d'écarter ceux associés aux valeurs propres les plus fortes.

En posant la matrice $A = U^T Z^T$, A permet la diagonalisation du numérateur et du dénominateur du critère de Fisher :

$$A S_w A^T = D_w, \quad A S_b A^T = I \quad (4.18)$$

En posant $\Phi = D_w^{-1/2} A$, nous obtenons la matrice de projection qui répond au critère de Fisher. La figure 4.7 montre la chaîne de l'algorithme de la DLDA.

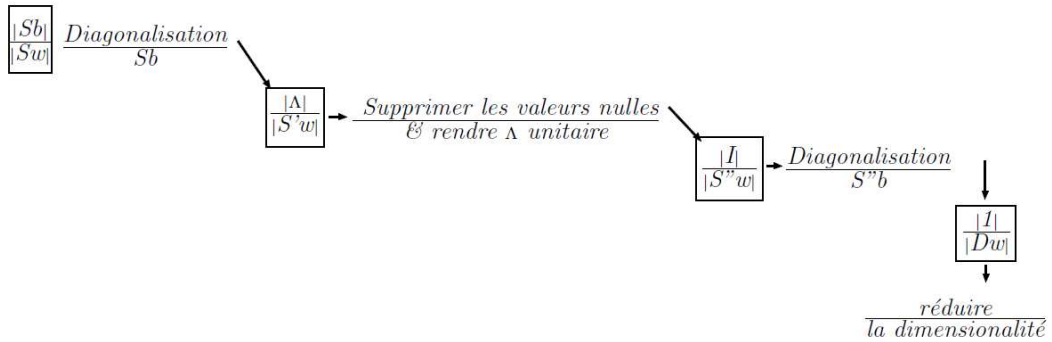


FIGURE 4.7: Diagramme de l'algorithme de la LDA directe (DLDA). La figure est tirée de [103].

4.5 Tests et comparaisons des algorithmes

Dans ce chapitre, nous avons vu les bases théoriques des algorithmes de reconnaissance de visage les plus utilisés. Ces algorithmes se basent sur l'approche globale par combinaison linéaire des composantes. Chacune de ces méthodes propose des solutions à certains problèmes rencontrés lors de la reconnaissance de

visage. Nous avons testé les différents algorithmes et les résultats qui en résulte confirment dans la plupart des cas les résultats de la littérature. Ces résultats présentent aussi quelques éclaircissements de comportement non attendu pour certains types d'expérience. Cette étude justifiera aussi le choix exclusif de l'algorithme de la DLDA comme algorithme d'étude dans la suite de cette thèse.

Bases de données, protocoles et résultats

Dans cette partie du chapitre, les résultats des 3 algorithmes, décrits dans la section précédente, sont présentés. Les bases de données de visage considérées sont les bases publiques FRGC et BANCA. Pour la base FRGC, nous avons choisi d'utiliser les deux protocoles d'évaluation : Expérience 1 et Expérience 4. Ces deux protocoles permettront d'étudier le comportement des différents algorithmes face aux problèmes des variabilités des conditions d'acquisition. Les mêmes profils d'expérience sont choisis pour la base BANCA. En considérant les protocoles Mc et P, les tests permettront de voir si les algorithmes gardent le même comportement avec un changement de la base d'évaluation. Le choix d'utiliser ces deux bases de données est justifié aussi par la différence sensible en terme de qualité car même si les profils des expériences des deux bases sont les mêmes, la résolution des images entre les deux bases diffère sensiblement. En effet, pour la base FRGC, la moyenne des distances entre les centres des yeux est de ~ 250 pixels alors que celle de la base BANCA est de ~ 100 pixels. Un autre point de divergence entre les expériences des deux bases est la validité statistique des résultats, à savoir que le nombre de tests de comparaison pour la base FRGC est beaucoup plus grand que celui de la base BANCA. Par conséquent, les résultats des expériences de la base FRGC devraient être généralisables sur la base BANCA.

Le point commun (ou le profil commun) entre ces différents protocoles est que la qualité de l'image de référence est considérée comme de bonne qualité car cette image est acquise dans des conditions contrôlées⁵ alors que les images de test sont de qualité variable.

Par la suite, nous considérons l'image du visage normalisée⁶ comme représentation du visage (ex : voir Figure 4.8). Afin de pouvoir analyser et comparer les

5. Le chapitre 3 section 3.4 détaille les différents protocoles pour chacune des deux bases.

6. Dans ce chapitre, nous utiliserons une normalisation géométrique avec une simple égalisation d'histogramme de la composante brillance de l'image comme mesure de l'illumination. Cette composante est extraite de l'espace colorimétrique (TSV). Une étude approfondie sur la normalisation est présentée dans le chapitre 5

différents algorithmes, nous avons choisi :

- de fixer la base d'apprentissage. Cette base servira pour le calcul des espaces réduits.
- d'utiliser les mêmes conditions expérimentales des protocoles d'évaluation.
- d'utiliser les mêmes conditions de pré-traitement des images des visages.

Ce choix permettra d'étudier la différence de comportements entre les différents algorithmes. Même étant dans les mêmes conditions de test, les différents algorithmes n'ont pas besoin des mêmes informations pour l'apprentissage. L'ACP qui est une méthode de réduction non supervisée n'a pas besoin d'information sur les classes (dans notre cas, les images d'un sujet forment une classe) alors que la LDA et la DLDA, qui sont des méthodes supervisées, en ont besoin.



FIGURE 4.8: Exemple d'image de brillance normalisée géométriquement

La base d'apprentissage **BA** (défini dans le chapitre 3) est une sous-partie de la base de développement de FRGC. Elle est constituée de 220 personnes avec 10 images de visage acquises dans des conditions contrôlées et 10 images acquises dans des conditions non contrôlées. Cette même base d'apprentissage sera également utilisée pour les protocoles de reconnaissance P et Mc de BANCA.

Résultats de reconnaissance sur les bases FRGC et BANCA pour les approches globales linéaires

Les résultats présentés par la suite corroborent, dans la plupart des cas, des résultats déjà constatés dans la littérature. On présente ici ces résultats afin de montrer les évolutions en performance des différents algorithmes par réduction d'espace et valider ainsi leur utilisation. Le choix de présenter les résultats sur la base FRGC se justifie par le fait que les résultats des protocoles de test sont statistiquement significatifs, comparés à toutes les bases publiques existantes. Le

protocole de l'expérience 1 comporte, par exemple, plus de 55.000.000 de tests inter-classes et près de 200 000 tests intra-classes pour 466 personnes.

Performances des algorithmes ACP, LDA et DLDA sur FRGC et BANCA :

Afin de mieux comprendre l'influence de la qualité des tests sur les différents algorithmes de reconnaissance de visage, basés sur la réduction d'espace, on propose d'observer les résultats de la reconnaissance de l'ACP, de la LDA et de la DLDA sur les 2 types d'expérience de FRGC (Expérience 1 et Expérience 4) ainsi que pour les deux protocoles P et Mc de BANCA. La base BA est utilisée comme base d'apprentissage des différents espaces de réduction (axes de projection) pour les différentes expériences.

Les courbes ROC de la figure de 4.9 présentent les résultats de l'expérience 1 et 4 de FRGC pour les 3 algorithmes cités (ACP, LDA et DLDA). Ceci nous permet de comparer les performances en fonction du caractère dégradé ou pas des données de test. Les courbes de la figure 4.10 représentent les courbes ROC pour les protocoles P et Mc de BANCA sur les deux groupes G1 et G2 (en utilisant la même base d'apprentissage BA) (voir chapitre 3 pour plus de détails sur les protocoles)⁷. Le tableau 4.3 résume les performances aux deux points de fonctionnement VR @0.1% de FAR et EER pour les deux bases selon les différents protocoles.

En comparant les résultats de reconnaissance pour les deux types de protocoles contrôlés et non contrôlés (Exp1 contre Exp4 pour FRGC et Mc contre P pour BANCA), nous constatons que les performances de l'ACP, de la LDA et de la DLDA sont bien meilleures (pour des références contrôlées) lorsque les images de test sont contrôlées que lorsqu'elles sont dégradées. En effet, par exemple le taux de VR @0.1% de FAR passe de 76% à 17% pour l'ACP entre l'expérience 1 et l'expérience 4. Cette tendance est bien vérifiée pour les 3 algorithmes sur les 2 type de protocoles.

Un deuxième point que nous pouvons constater est l'amélioration des performances obtenues par la DLDA par rapport aux deux autres méthodes. Ce résultat est visible dans tous les tests. Ainsi pour l'expérience 1 de FRGC nous observons une différence significative entre la DLDA (83% de VR) et l'ACP et la LDA (respectivement 76% et 58%). Le même comportement peut être observé pour tous les tests.

7. Notons que je présente ici les résultats en terme de courbes ROC et non pas en WER ("*Weighted Error Rate*")

Finalement nous notons aussi une différence de performances entre la LDA et l'ACP. Les performances de l'ACP sont supérieurs à ceux de la LDA pour tous les tests sauf pour l'expérience 4 de FRGC.

Les performances de l'ACP sont meilleures lorsque le test est fait en conditions contrôlées. L'espace de réduction dans l'expérience 1 n'est adapté ni à la LDA ni à la PCA, car on apprend cet espace sur des images contrôlées et des images non contrôlées (5 images acquises en conditions contrôlées et 5 images dans des conditions non contrôlées) et dans ce cas on a pas assez d'images des conditions contrôlées [52]. L'ACP reste insensible à ce problème que la LDA, parce qu'elle tend à maximiser la variance des échantillons et non à maximiser la séparation entre les classes comme pour la LDA. Les données étant mal définis dans l'espace de projection l'erreur induite par la LDA est donc plus grande que l'ACP. Par contre dans le contexte de l'expérience 4, le caractère discriminant entre les classes amené par la LDA montre toute son efficacité.

Dans le cas des expériences sur BANCA, la base d'apprentissage des espaces de projection (BA de FRGC) ne correspond pas aux personnes de la base de test. On retrouve les mêmes résultats que dans [52] à savoir que dans ce cas l'ACP est meilleure que la LDA et ceci indépendamment des conditions de tests contrôlés ou pas.

Résultats de la base FRGC

Algorithme	Expérience 1 (VR/EER)	Expérience 4(VR/EER)
ACP	76.37% \pm 0.18] / 5.23 \pm 0.09]%	17.14% \pm 0.23] / 26.40 \pm 0.27]
LDA	58.90% \pm 0.21] / 7.73 \pm 0.11]%	25.29% \pm 0.27] / 15.51 \pm 0.22]
DLDA	83.19% \pm 0.15] / 2.93 \pm 0.07]%	33.17% \pm 0.29] / 10.69 \pm 0.19]

Résultats de la base BANCA pour G1

Algorithme	Protocole Mc (VR/EER)	Protocole P(VR/EER)
ACP	64.3% \pm 4.5%] / 12.1% \pm 3.0%]	29.9% \pm 2.6%] / 19.2% \pm 2.3%]
LDA	50% \pm 4.9%] / 15.3% \pm 3.4%]	22.73% \pm 2.4%] / 24.7% \pm 2.4%]
DLDA	66.4% \pm 4.6%] / 11.7% \pm 3.0%]	29.9% \pm 2.6%] / 18.6% \pm 2.2%]

Résultats de la base BANCA pour G2

Algorithme	Protocole Mc (VR/EER)	Protocole P(VR/EER)
ACP	81.5% \pm 3.8%] / 4.0 \pm 2.0%]	34.1% \pm 2.7%] / 21.7% \pm 2.3%]
LDA	55.1% \pm 4.9%] / 13.5 \pm 3.4%]	25.0% \pm 2.5%] / 22.9% \pm 2.4%]
DLDA	88.9% \pm 3.1%] / 4.3 \pm 1.9%]	42.5% \pm 2.8%] / 20.1% \pm 2.2%]

TABLE 4.3: Résumé des résultats aux points de fonctionnement (EER et VR@0.1% de FAR) des 3 algorithmes (ACP, LDA, DLDA) pour les deux types d'expérience sur les bases FRGC et BANCA

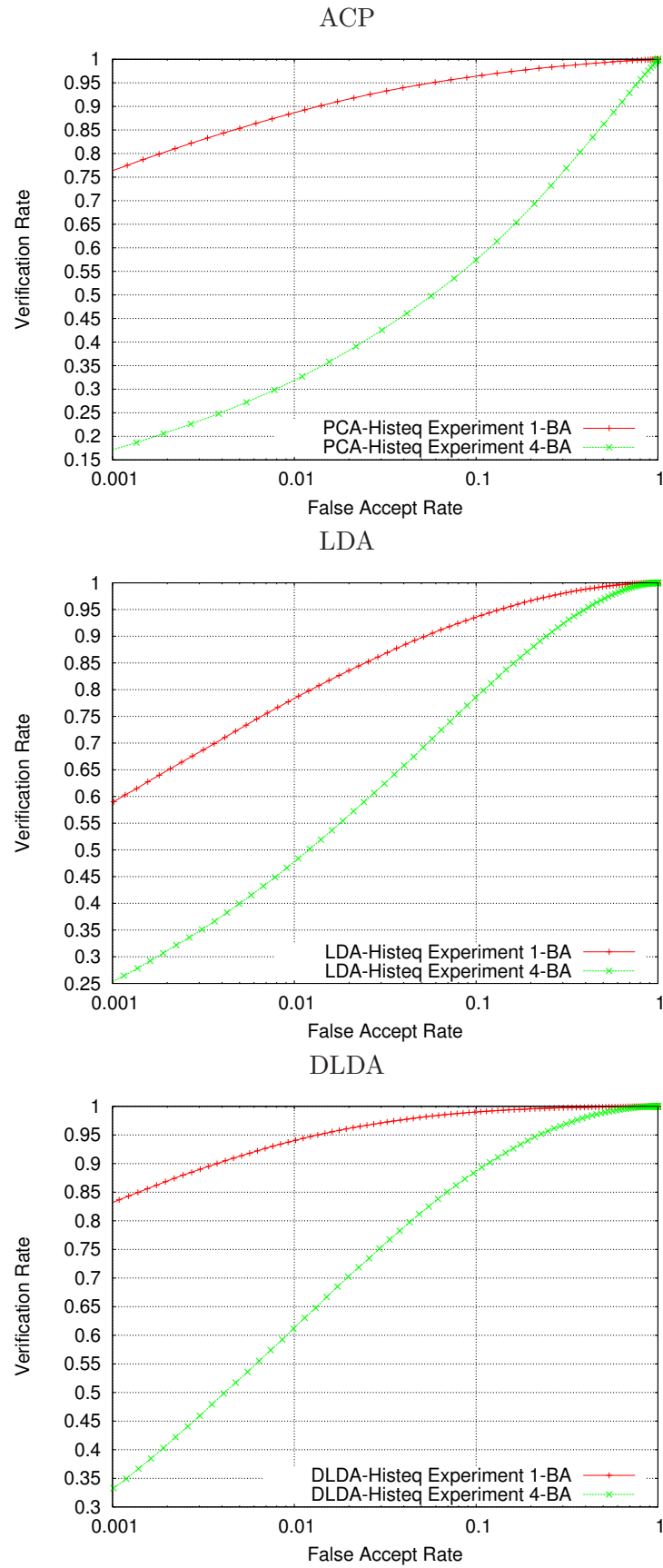


FIGURE 4.9: Courbes ROC des algorithmes ACP, LDA et DLDA pour les expériences 1 et 4 de FRGC : avec égalisation d'histogramme, base d'apprentissage BA

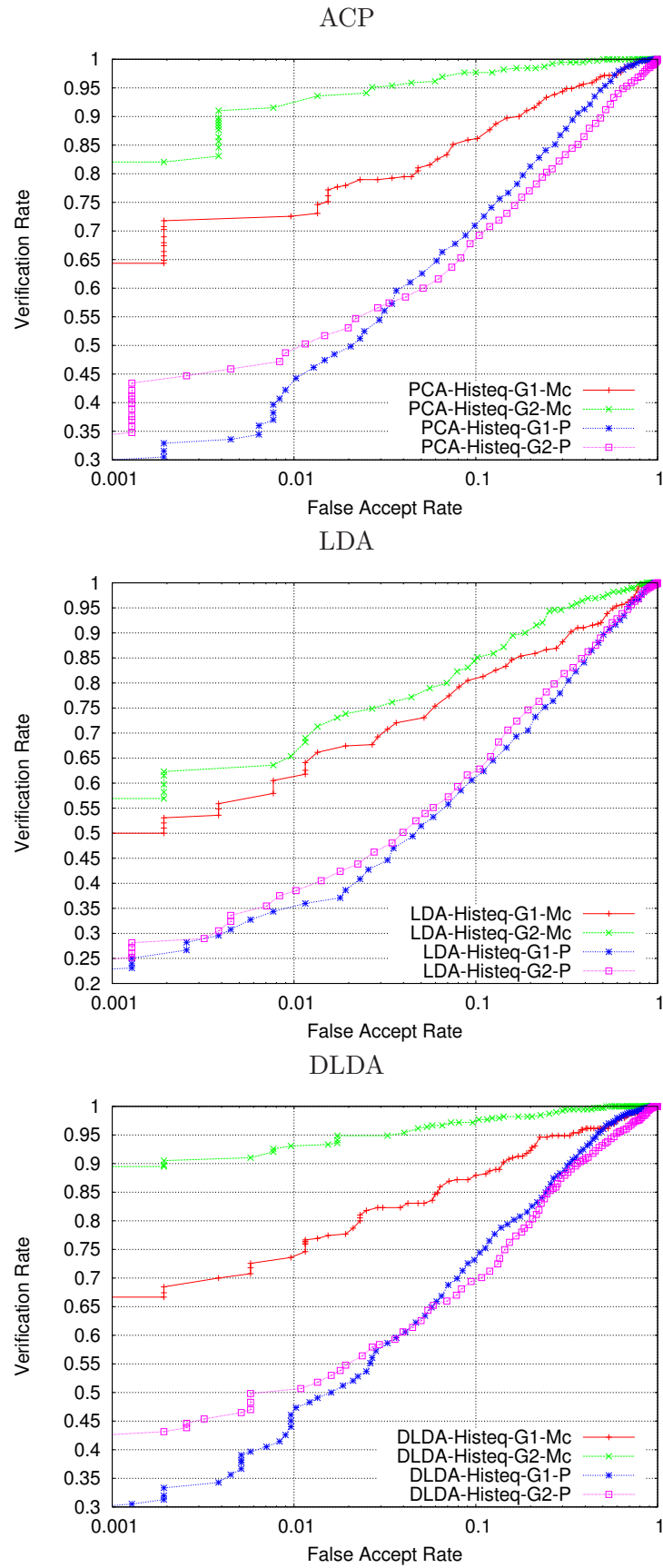


FIGURE 4.10: Courbes ROC des algorithmes ACP, LDA et DLDA pour les expériences Mc et P de la base BANCA : avec égalisation d'histogramme, base d'apprentissage BA

Comparaison des algorithmes ACP, LDA et DLDA sur la base FRGC :

En analysant les résultats de l'expérience 1 (Figure 4.11), nous remarquons que les résultats de la LDA se dégradent par rapport à ceux de l'ACP. Ceci s'explique par le fait que l'ACP perd une partie de l'information discriminante entre les classes, qui est plus importante lorsque les conditions d'acquisition sont contrôlées (absence de bruit) que lorsque ces conditions sont non contrôlées. Les résultats de la DLDA sont bien meilleurs que ceux de l'ACP ou ceux de la LDA. Cela confirme que la réduction par ACP avant le calcul des axes de projection par LDA supprime une quantité importante d'information discriminante. Les performances de l'ACP sont meilleures que la LDA dans cette expérience en conditions contrôlées. L'analyse de la courbe Roc de la figure 4.12 nous donne des résultats comparables.

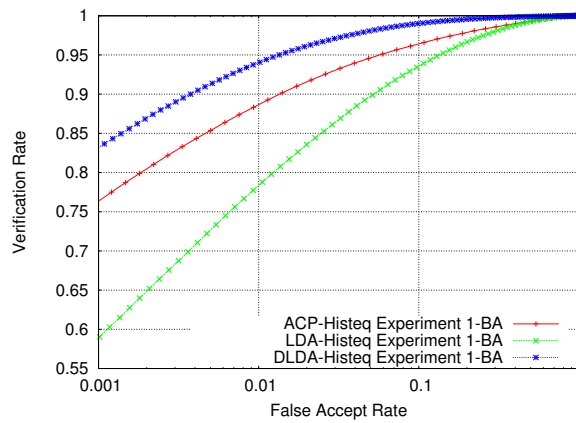


FIGURE 4.11: Comparaison des performances pour les algorithmes ACP, LDA et DLDA pour l'expérience 1 sur la base FRGC

Comme pour l'expérience 1, les performances de la DLDA sont meilleures pour l'expérience 4 (figure 4.12) que celle de l'ACP et de la LDA ce qui démontre la robustesse de cette méthode lorsque les conditions d'acquisition sont difficiles. Par contre les performances de la LDA sont meilleures que l'ACP dans cette expérience.

Ces résultats sont en accord avec d'autres travaux publiés qui comparent LDA et PCA [52].

Évolution des performances de la DLDA en fonction du type de base d'apprentissage :

Un autre point connu pour les méthodes de réduction de l'espace, mais souvent omis dans la littérature, est l'influence de la base d'apprentissage sur les perfor-

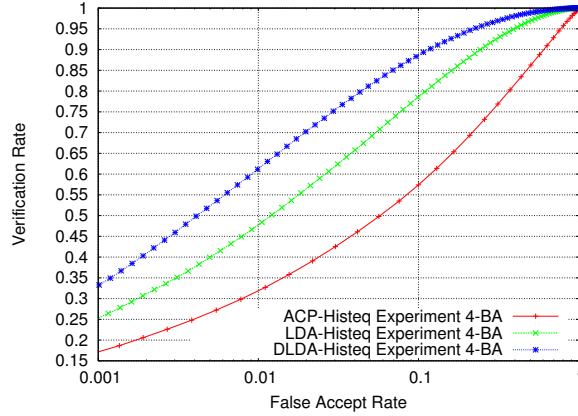


FIGURE 4.12: Comparaison des performances pour les algorithmes ACP, LDA et DLDA pour l'expérience 4 sur la base FRGC

mances de l'évaluation. On présente ici les performances de l'algorithme DLDA en fonction du type de base d'apprentissage et du protocole de test.

Afin d'étudier l'influence de la base d'apprentissage, on considère les protocoles suivants : Expérience 1 (image de référence contrôlée et image de test contrôlée) et Expérience 4 (image de référence contrôlée et image de test non contrôlée).

Nous construisons plusieurs espaces de réduction à partir de sous-bases de la base d'apprentissage BA (220 personnes et 10 images de type contrôlé et 10 images de type non contrôlé par personne).

Deux ensembles d'espace sont alors construits. Dans le premier ensemble, nous fixons le nombre d'images acquises en conditions contrôlées à $C = 10$ et nous varions le nombre d'images acquises dans les conditions non contrôlées à $NC = 1, 2, \dots, 10$. Dans le deuxième ensemble, nous varions le nombre d'images acquises en conditions contrôlées ainsi que le nombre d'images acquises dans les conditions non contrôlées tel que $C + NC = 10$.

Les figures 4.13 et 4.14 montrent les performances de l'algorithme DLDA en fonction du type de l'espace d'apprentissage.

Nous pouvons constater que pour l'expérience 1 (images référence et de test acquises dans des conditions contrôlées), un apport de bruit dans la base d'apprentissage dégrade les performances du système. Pour l'expérience 4 (images référence acquises en conditions contrôlées et images de test acquises en conditions non contrôlées), nous pouvons remarquer qu'un mélange des deux types d'images dans la base d'apprentissage est très important afin d'améliorer les résultats.

Ces résultats montrent une forte dépendance entre la qualité des images de

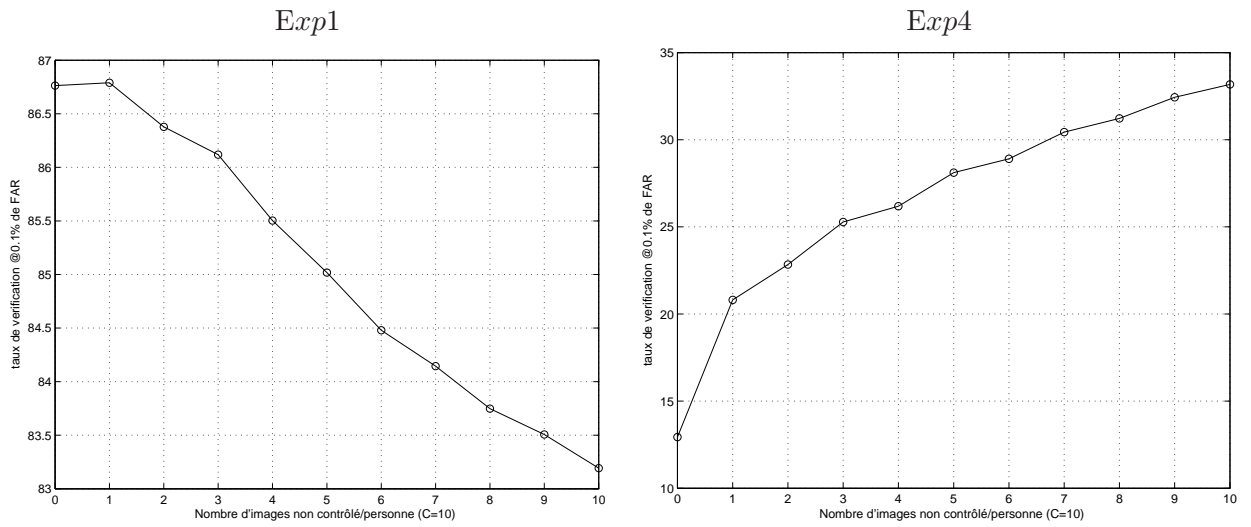


FIGURE 4.13: Variation des performances de la DLDA en fonction de la variation du nombre d'images acquises en conditions non contrôlées

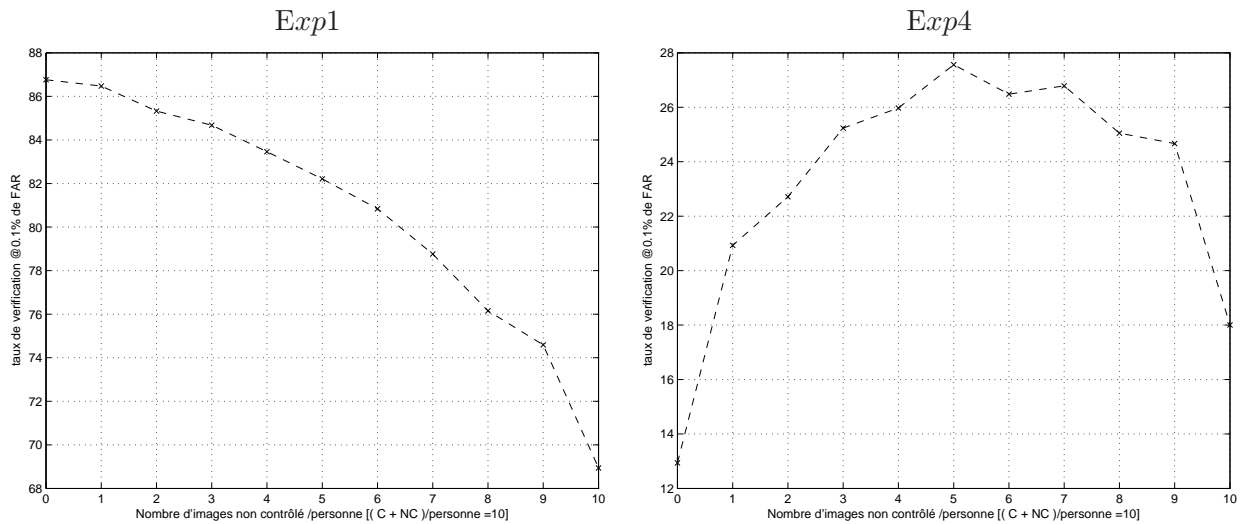


FIGURE 4.14: Variation des performances de la DLDA en fonction de la variation proportionnelle du nombre d'images acquises en conditions contrôlées et non contrôlées

test, la qualité des images constituant la base d'apprentissage et les performances du système. Cette constatation est très importante pour la suite de nos études et cette dépendance sera étudiée plus en détails dans le dernier chapitre pour le choix de la stratégie de reconnaissance.

4.6 Conclusion

Dans ce chapitre, nous avons étudié trois algorithmes de reconnaissance de visage permettant de réaliser des réductions linéaire d'espace. Nous avons exposé, dans la première partie, la théorie de ces différents algorithmes et nous avons expliqué les raisons de l'efficacité de ce type d'approches pour la reconnaissance de visage. Nous nous sommes concentrés sur plusieurs problèmes que rencontrent les approches par réduction d'espace telle que le choix des bases de données d'apprentissage ou le problème de la sélection de la dimensionnalité de l'espace de réduction. Nous avons étudié l'impact du choix des bases d'apprentissage sur les performances des algorithmes suivant la nature des images (contrôlées ou dégradées).

Notre choix s'est finalement porté sur l'algorithme DLDA qui a démontré de meilleures performances que l'ACP et la LDA, notamment sur les données dégradées.

Chapitre 5

Normalisation de l'illumination

5.1 Introduction

La normalisation des images de visages est une étape très importante pour les algorithmes de reconnaissance. Généralement, les algorithmes se basant sur les approches par points caractéristiques ont recours à des normalisations photométriques alors que les approches basées sur la réduction d'espace (les approches globales) ont besoin en plus d'une normalisation géométrique.

Dans ce chapitre, nous nous concentrerons sur certaines normalisations photométriques des images du visage et nous présenterons et analyserons les résultats de reconnaissance par l'approche DLDA basée sur ces méthodes.

Les méthodes étudiées dans ce chapitre sont : l'égalisation d'histogramme, la correction Gamma, la méthode "multiretinex" et le lissage anisotropique. Nous avons choisi d'utiliser ces méthodes car elles n'ont besoin que d'un seul exemple de l'image pour être appliquées.

Les deux bases de données FRGC et BANCA ont été utilisées afin de pouvoir comparer les performances de l'algorithme de reconnaissance DLDA (voir chapitre précédent), suivant les différentes méthodes de normalisation.

À la fin de ce chapitre, nous présentons une nouvelle méthode de normalisation de l'illumination qui améliore de façon significative les résultats de la reconnaissance dans les conditions dégradées.

5.2 Normalisation géométrique

En utilisant un algorithme de reconnaissance se basant sur la réduction de l'espace, nous ne pouvons pas négliger un point très important qui est la normalisation géométrique des images de visage. Cette normalisation géométrique consiste à extraire la zone du visage de l'image originale, ensuite une rotation du visage est effectuée afin d'aligner l'axe des yeux avec l'axe horizontal. Enfin, une réduction proportionnelle à la distance entre les centres des deux yeux est appliquée. On obtient alors une image de visage dont la distance entre les centres des yeux est fixe.

Les dimensions de l'image du visage sont calculées à partir de la distance à obtenir entre les centres des deux yeux.

$$heightNorm = \lambda * EyeDistance$$

$$WidthNorm = \gamma * EyeDistance$$

avec *heightNorm* et *WidthNorm* la hauteur et la largeur de l'image du visage, *EyeDistance* la distance entre les centres des yeux de l'image normalisée. Nous fixons également, lors de cette étape, la position du centre de la bouche dans l'image normalisée afin d'avoir une bonne normalisation verticale et d'être sûr que les différents composants du visage (yeux, nez et bouche) soient situés dans les mêmes positions pour tous les visages. L'exemple de la figure 5.1 illustre une normalisation géométrique d'un visage de la base FRGC.



FIGURE 5.1: Exemple d'une normalisation géométrique de visage de la base FRGC

Un accroissement de la dynamique est aussi appliqué à l'image normalisée. Cet accroissement est basé sur un centre-réduction de l'histogramme de l'image pour aboutir à des images avec les mêmes plages de répartition des niveaux de gris ainsi qu'un alignement des moyennes de ces niveaux.

Evolution des performances de la DLDA en fonction de la normalisation géométrique

La figure 5.2 représente la variation du taux de EER de l'algorithme de la DLDA en fonction de la distance séparant les centres des yeux. Ces expériences sont les résultats du protocole Expérience 1 de la base FRGC. La base d'apprentissage utilisée est la base BA (un sous-ensemble de la partie développement de FRGC, voir chapitre précédent).

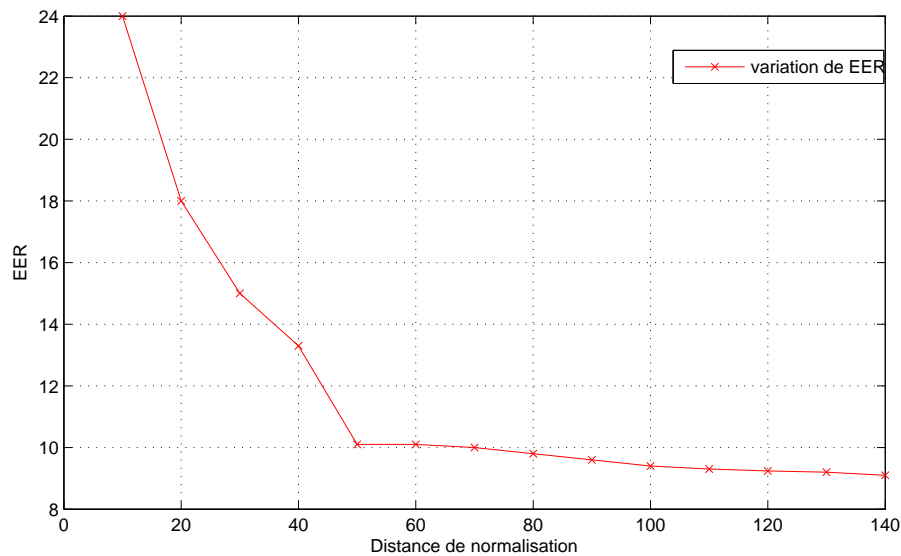


FIGURE 5.2: Variation du taux de EER de l'algorithme de la DLDA en fonction de la normalisation géométrique sur la base FRGC pour l'expérience 1

La courbe de variation 5.2 montre que les résultats pour les images normalisées de petite taille (<20 pixels) ne sont pas très satisfaisants. Une distance de normalisation de > 25 pixels améliore sensiblement les résultats, nous passons d'un EER de $\sim 24\%$ à 15% pour la distance de normalisation de 30 pixels. Les performances de la DLDA se stabilisent à partir d'une distance de normalisation supérieure à 45 pixels (12% de EER). Le EER atteint une valeur minimum de 9.1% pour une distance de normalisation de 150 pixels. Nous avons choisi une distance de normalisation de 50 pixels entre les centres des yeux. Ce choix nous semble bon pour équilibrer le rapport performance et complexité du système.

NB : Dans la suite de la thèse, le terme normalisation indiquera implicitement une normalisation géométrique avec accroissement de la dynamique.

5.3 Extraction de l'intensité du pixel de l'espace couleur

La plupart des images des bases de données publiques de visage sont en couleurs (BANCA, XM2VTS, FRGCv1, FRGCv2, FERET..). Cette quantification de la couleur est faite dans le système RVB (Rouge, Vert et Bleu). Les algorithmes de reconnaissance de visages utilisent, généralement, le niveau de gris comme mesure d'intensité de cette couleur. L'utilisation de la couleur pose un certain nombre de problèmes à savoir que la couleur dépend fortement de la calibration du capteur. Les niveaux des couleurs ne sont pas perçus de la même manière pour un même objet sous différentes conditions de l'illumination et la correction de cette variation est très difficile à mettre en oeuvre. En utilisant la mesure de l'intensité d'illumination, nous évitons de travailler avec les composantes couleurs et ainsi de chercher l'espace couleurs optimal qui représente les visages. La mesure d'intensité peut être obtenue par plusieurs transformations du système RVB [19], je présente ici les trois transformations les plus utilisées dans la littérature :

- Par moyenne des 3 composantes (Rouge, Vert et Bleu) de l'image ;
- Par passage à un autre espace de couleur TSV (Teinte, Saturation, Valeur) (voir figure 5.3(b)) ;
- Par passage à l'espace de couleur TSL (Teinte, Saturation, Luminance) (voir figure 5.3(c)).

La première solution présente un grand inconvénient, car même si l'aspect visuel de l'image reste correct, cette transformation linéaire ne prend pas en compte la prépondérance d'une composante couleur par rapport aux autres et le contraste de l'image est alors réduit.

La deuxième et la troisième solution expriment mieux, en général, le niveau d'intensité des couleurs que le système RGB. Il ne faut pas confondre le modèle TSL avec le modèle TSV. En TSV, les couleurs pures et le blanc ont la même "valeur". En TSL, seul le blanc est considéré comme une lumière de 100% [25].

Passage de RVB à TSV et TSL

Le passage du système de couleurs RVB au système TSV ou TSL se fait par une fonction de transfert non linéaire. La différence entre les systèmes de couleurs TSV et TSL, qui nous intéressent, réside principalement dans la définition des composantes "Valeur" et "Luminance".

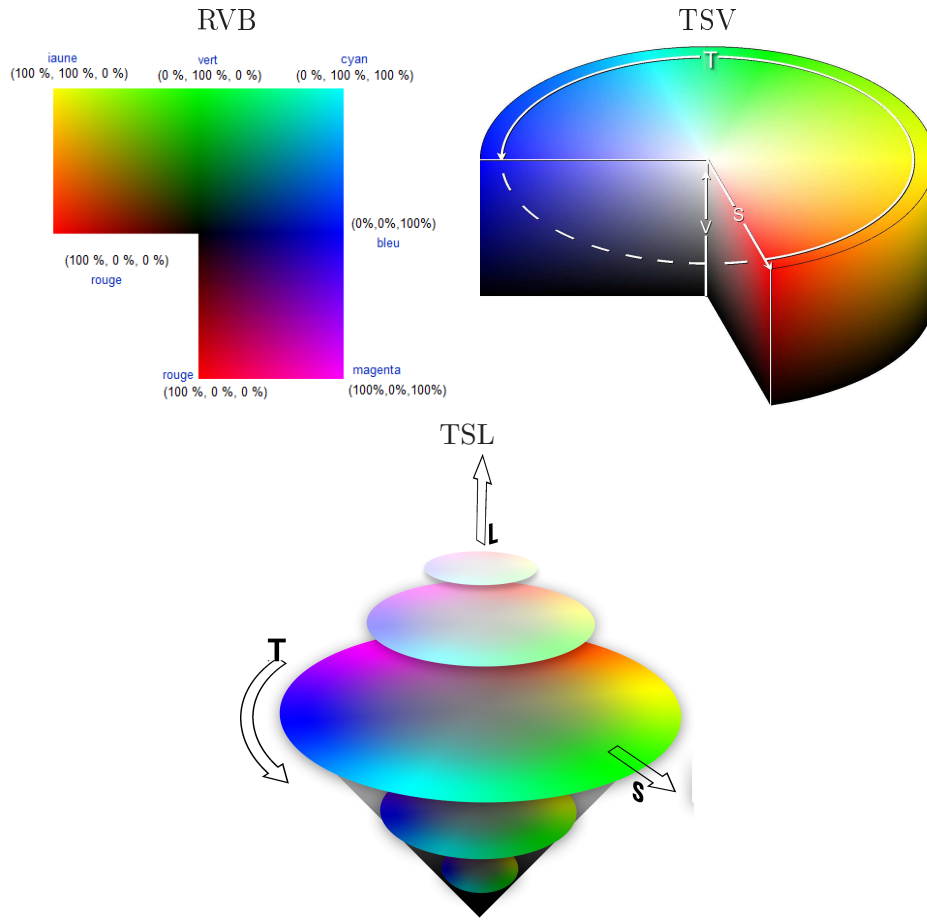


FIGURE 5.3: Différents systèmes colorimétriques : (a) plan du système de couleurs RVB, (b) cylindre du système de couleurs TSV, (c) cône du système de couleurs TSL

Soient \min et \max , respectivement, les valeurs minimales et maximales des composantes R, V et B ($\min = \text{minimum}(R, V, B)$ et $\max = \text{maximum}(R, V, B)$) et soient x et y les coordonnées d'un pixel de l'image.

Les transformations des espaces RVB vers TSV et TSL (Teinte (T), Saturation

(S), Valeur (V) et Luminance (L)) sont données par les équations suivantes :

$$T(x, y) = \begin{cases} \text{non défini, si } \min = \max \\ 60 * \frac{V - B}{\max - \min}, \text{ si } \max = R \text{ et } V \geq B \\ 60 * \frac{\max - \min}{V - B} + 360, \text{ si } \max = R \text{ et } V < B \\ 60 * \frac{\max - \min}{B - R} + 120, \text{ si } \max = V \\ 60 * \frac{R - V}{\max - \min} + 240, \text{ si } \max = B \end{cases}$$

$$S(x, y) = \begin{cases} 0, \text{ si } \max = 0 \\ 1 - \frac{\min}{\max}, \text{ sinon} \end{cases}$$

Pour le TSV :

$$V(x, y) = \max$$

Pour le TSL :

$$L(x, y) = \frac{\max + \min}{2}$$

pour plus de détails sur ces transformations, voir [25].

Ayant besoin de l'intensité de la couleur, nous allons nous intéresser aux composantes "Valeur" et "Luminance", correspondant respectivement aux systèmes TSV et TSL. Nous montrerons dans la dernière partie de ce chapitre que pour l'application de reconnaissance des visages, la composante "Valeur" de l'espace TSV donne de meilleurs résultats que la composante "Luminance" du système colorimétrique TSL.

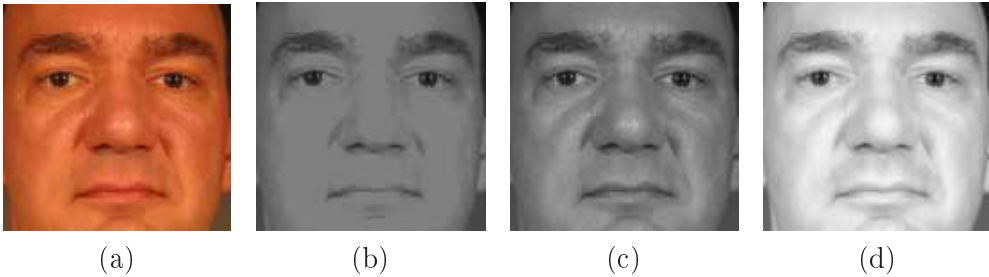


FIGURE 5.4: Différentes transformations de l'image couleur d'un exemple de la base FRGC (a) image couleur, (b) niveaux de gris par moyenne des 3 composantes RVB, (c) niveaux de gris par TSL, (d) niveaux de gris par TSV.

5.4 Normalisation de l'illumination

Dans le domaine de la reconnaissance par le visage, un certain nombre de méthodes de normalisation de l'illumination ont été présentées. Ces méthodes peuvent être classées en 2 grandes catégories, les méthodes basées sur la modification de la dynamique (égalisation d'histogramme, égalisation d'histogramme adaptative, transformation linéaire ou non linéaire de l'histogramme ...) [35] et des méthodes basées sur l'estimation de la réflectance de l'image (Retinex, MultiScale-Retinex, filtrage homomorphique, méthodes basées sur l'isotropie de la luminance et d'autres sur l'anisotropie de la réflectance) [26, 94].

Les 4 normalisations de l'état de l'art présentées dans ce chapitre sont les plus utilisées dans la reconnaissance du visage et font partie de ces deux catégories. Il s'agit de :

- **L'égalisation d'histogramme** : Le but de cette normalisation est de répartir uniformément les niveaux de gris de l'image afin d'améliorer son contraste.
- **La correction Gamma** [] : cette normalisation permet de réduire l'écart entre les zones sombres et claires du visage par une transformation non linéaire des niveaux de gris.
- **Le MultiScale-Retinex** [94] : L'objectif de cette méthode est d'estimer la partie luminance de l'image pour en déduire la réflectance.
- **Le lissage anisotropique** [26] : cette approche estime aussi la luminance de l'image mais en se basant sur l'anisotropie de la réflectance.

Nous présentons à la fin de ce chapitre une méthode originale, qui se base sur l'approche par lissage anisotropique, elle permet de corriger les défauts de cette méthode et de s'affranchir des problèmes d'illumination en préservant l'information texture.

5.4.1 Égalisation d'histogramme

Cette normalisation appartient à la catégorie des méthodes basées sur l'amélioration de la dynamique de l'image. Le but de l'égalisation est d'harmoniser la répartition des niveaux de gris de l'image. Elle permet de tendre vers un même nombre de pixels pour chacun des niveaux de l'histogramme. Cette opération vise à augmenter les nuances dans l'image et donc son contraste [Figure 5.5].

Plus concrètement, si n_i est le nombre de pixels à un niveau i de gris, la probabilité

qu'un pixel x de l'image ait un niveau i est :

$$p(x_i) = \frac{n_i}{n}, i \in 0, \dots, L \quad (5.1)$$

avec n le nombre total des pixels de l'image et L le nombre des niveaux de gris. La probabilité p représente dans ce cas l'histogramme de l'image normalisée à $[0, 1]$. Soit c la distribution cumulative de l'histogramme normalisé p , elle est donnée par :

$$c(i) = \sum_{j=0}^i p(x_j) \quad (5.2)$$

L'idée est de trouver une transformation $y = T(x)$ qui, pour chaque niveau x de l'image, produira un niveau y de telle façon que la distribution cumulative des différents niveaux de l'image transformée soit linéaire. Cette fonction est définie par $y_i = T(x_i) = L \frac{c(i)}{n}$ avec L le niveau maximal de gris.

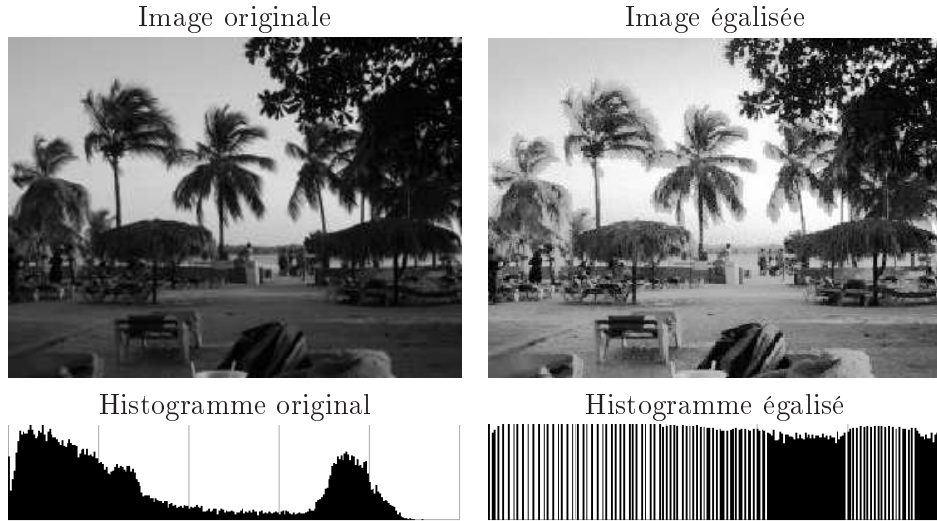


FIGURE 5.5: Exemple d'égalisation d'histogramme

5.4.2 Correction Gamma

Cette normalisation appartient à la catégorie des méthodes basées sur la modification de la dynamique de l'image. A l'histogramme original de l'image, on applique une transformation non linéaire dans le but de corriger des grandes variations de contraste. Les zones les plus sombres seront alors rehaussées alors que les zones brillantes seront réduites :

$$y = \lambda * x^{\frac{1}{\gamma}}, x \in 0, \dots, L \quad (5.3)$$

avec x le niveau de gris, L le maximum des niveaux de gris, γ le facteur de correction (en général, ce facteur dépend de l'application) et λ une constante de gain. Pour nos tests, la constante de gain λ a été fixée empiriquement à 1 et le facteur de correction γ a été optimisé sur la base de développement de FRGC. Les exemples (1.c) et (2.c) de la figure 5.8 illustrent une correction gamma de l'image du visage.

Les deux méthodes MultiRetinex et lissage anisotropique qui seront présentées à la suite, se basent sur la séparation des deux composantes de l'image : la luminance et la réflectance. La modélisation de l'image par combinaison de la réflectance et de la luminance a été proposée par Barrow et Tenenbaum en 1978 [5]. Ceci dit, en raison des différents facteurs qui peuvent entrer en jeu pour la construction de l'image (l'illumination de l'objet, la géométrie de la scène acquise, les paramètres de la caméra...), une telle modélisation reste très difficile à mettre en œuvre.

En 1999, Laszlo [90] a proposé un modèle génératif de l'image, basé sur la combinaison de l'intégrale de Fredholm et d'une modélisation des paramètres de la caméra. Cette modélisation reste assez complexe à mettre en œuvre¹.

La modélisation par réflexion diffuse de l'image [90] reste la plus utilisée :

$$I(x, y) = L(x, y) * R(x, y) * \cos_{\theta}(x, y) \quad (5.4)$$

avec x et y les coordonnées d'un pixel de l'image, $I(x, y)$ le niveau de gris du pixel, $L(x, y)$ l'amplitude de l'illumination en ce point (dépendante de la source), $R(x, y)$ la réflectance du point ou le coefficient de réflectance (dépendante de la nature intrinsèque de la surface de l'objet) et $\cos_{\theta}(x, y)$ le cosinus de l'angle entre le rayon lumineux incident et la surface normale au point de l'objet [Figure 5.6].

En traitement d'image, cette modélisation a été encore plus simplifiée en intégrant $\cos_{\theta}(x, y)$ dans la composante $L(x, y)$. Le modèle final devient alors :

$$I(x, y) = L(x, y) * R(x, y) \quad (5.5)$$

Cette modélisation de l'image est loin d'être parfaite, car elle ne prend pas en compte ni les problèmes liés à la géométrie de l'objet (présence de surfaces qui peuvent créer des ombres sur l'objet...), ni les bruits, ni les facteurs externes à la formation de l'image. L'avantage de cette modélisation simple est de pouvoir estimer la réflectance d'un objet à partir d'une approximation de sa luminance.

1. Le document [90] présente un grand nombre d'autres modélisations de l'image

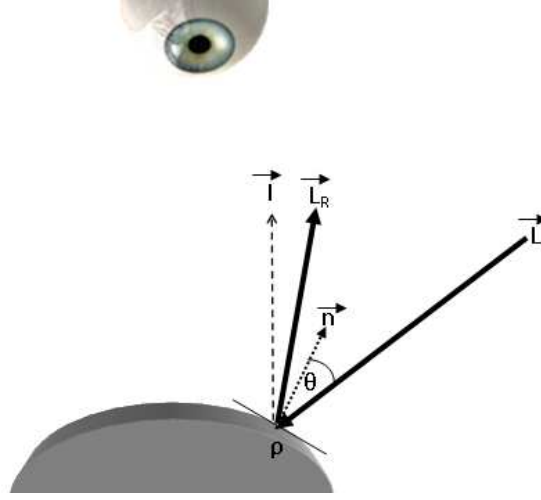


FIGURE 5.6: Schématisation de la réflexion diffuse. C'est la quantité de lumière I reçu par l'oeil (ou par un capteur). Pour un objet lambertien, c'est la projection de \vec{L}_R sur l'axe vertical, avec \vec{L}_R la réflexion du vecteur luminance L sur la surface de l'objet pondéré par son coefficient de réflectance ρ .

L'estimation de la réflectance est très importante, car nous avons ici une possibilité de caractériser un objet indépendamment des problèmes liés à l'illumination.

5.4.3 MultiScale-Retinex

La luminance peut être considérée comme une version lissée de l'image originale. En effectuant une convolution de l'image originale par un filtre gaussien, nous obtenons une estimation de la luminance. Proposée par Edwin H. Land en 1937, cette méthode est appelée "*retinex*" ou "*Single Scale Retinex*". Cette idée n'a pas vraiment de preuves physique mais elle a le mérite de donner des résultats assez exceptionnels.

$$L(x, y) = I(x, y) \otimes G(x, y) \quad (5.6)$$

La méthode MSR (*MutliScale-Retinex*) est un dérivé de la méthode "*Single Scale Retinex*". En 1997, Rahman [94] a proposé une estimation de la luminance comme combinaison pondérée d'images filtrées de l'image originale. Les filtrages se font par des noyaux gaussiens avec différentes variances (dans un sens, c'est une analyse multi-résolution, d'où le terme "multi-scale").

$$L(x, y) = \sum_{j=1}^S w_s(I(x, y) \otimes G_s(x, y)) \quad (5.7)$$

avec w_s un coefficient de pondération et G_s un noyau gaussien avec un écart type σ_s .

Une fois la luminance déterminée, on peut déduire la réflectance de l'objet en divisant l'image des niveaux du gris par l'image luminance.

$$R(x, y) = \frac{I(x, y)}{L(x, y)} \quad (5.8)$$

5.4.4 Lissage anisotropique

Comme pour la méthode multiRetinex l'objectif de cette approche est d'extraire la réflectance après modélisation de la luminance. La luminance L est aussi considérée, pour cette approche, comme une fonction lissée de l'image originale. Cette fonction peut être construite de façon à avoir une image similaire à l'originale en imposant une contrainte sur la fonction de lissage. La luminance est construite alors en minimisant la fonctionnelle J :

$$J(L) = \int_y \int_x (L - I)^2 dx dy + c \int_y \int_x (L_x^2 + L_y^2) dx dy \quad (5.9)$$

Le premier terme de la fonctionnelle J modélise la ressemblance entre I et L et le deuxième terme modélise la contrainte de lissage avec c le coefficient de contrôle de la contrainte de lissage et L_x et L_y les dérivés de L dans les deux directions x et y . Le problème 5.9 correspond à un problème de minimisation de fonctionnelle qui se résout par l'équation d'Euler-Lagrange qui a comme solution :

$$(L - I) + c(L_x + L_y = 0) \quad (5.10)$$

En discrétisant cette solution sur le voisinage de chaque pixel [Figure 5.7], on aboutit à la solution discrète suivante :

$$I_{i,j} = L_{i,j} + c(\nabla_N L_{i,j} + \nabla_S L_{i,j} + \nabla_E L_{i,j} + \nabla_W L_{i,j}) \quad (5.11)$$

avec :

$$\begin{aligned} \nabla_N L_{i,j} &= L_{i,j} - L_{i-1,j} \\ \nabla_S L_{i,j} &= L_{i,j} - L_{i+1,j} \\ \nabla_E L_{i,j} &= L_{i,j} - L_{i,j+1} \\ \nabla_W L_{i,j} &= L_{i,j} - L_{i,j-1} \end{aligned}$$

Gross et Brajovic [26] ont proposé de généraliser la fonctionnelle J de l'équation 5.9 en ajoutant un poids $\rho(i, j)$ au premier terme pour modéliser au mieux

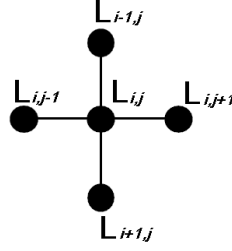


FIGURE 5.7: Discrétisation au voisinage d'un pixel

la perception réelle de l'œil humain (d'après les travaux de Weber en vision et neurobiologie).

$$J(L) = \int_y \int_x \rho(L - I)^2 dx dy + c \int_y \int_x (L_x^2 + L_y^2) dx dy \quad (5.12)$$

La solution pour l'équation 5.12 est donnée par

$$I = L + \frac{c}{\rho}(L_x + L_y) \quad (5.13)$$

En discrétisant cette solution sur le voisinage de chaque pixel, on aboutit à :

$$I_{i,j} = L_{i,j} + c \left(\frac{1}{\rho_N} \nabla_N L_{i,j} + \frac{1}{\rho_S} \nabla_S L_{i,j} + \frac{1}{\rho_E} \nabla_E L_{i,j} + \frac{1}{\rho_W} \nabla_W L_{i,j} \right) \quad (5.14)$$

Avec :

$$\begin{aligned} \rho_N &= \frac{|I_{i,j} - I_{i-1,j}|}{\min(I_{i,j}, I_{i-1,j})} \\ \rho_S &= \frac{|I_{i,j} - I_{i+1,j}|}{\min(I_{i,j}, I_{i+1,j})} \\ \rho_E &= \frac{|I_{i,j} - I_{i,j+1}|}{\min(I_{i,j}, I_{i,j+1})} \\ \rho_W &= \frac{|I_{i,j} - I_{i,j-1}|}{\min(I_{i,j}, I_{i,j-1})} \end{aligned}$$

Ce coefficient est appelé "coefficient de weber".

Une fois la luminance déterminée, on peut déduire la réflectance, comme pour la méthode MultiRetinex, en divisant l'image des niveaux du gris par l'image luminance.

$$R(x, y) = \frac{I(x, y)}{L(x, y)} \quad (5.15)$$

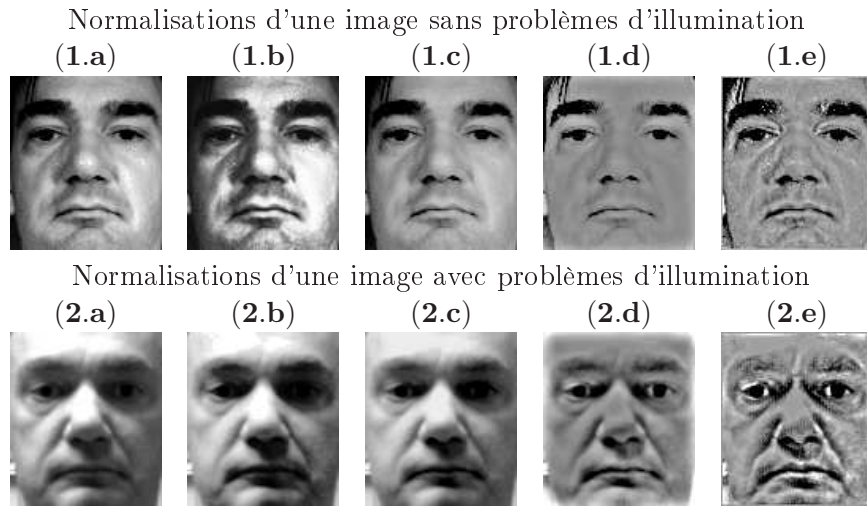


FIGURE 5.8: Normalisation du visage : (x.a) géométrique, (x.b) égalisation d'histogramme, (x.c) Gamma, (x.d) MultiScale-Retinex et (x.e) lissage anisotropique

5.4.5 Nouvelle méthode de correction de l'image par transformation de l'illumination

Comme présenté dans la première partie de ce chapitre, les deux approches utilisées dans la reconnaissance par le visage se basent sur la correction de la dynamique ou sur la séparation de la luminance et de la réflectance de l'image. Chacune de ces approches présente des avantages et des inconvénients. Les normalisations utilisant la correction de la dynamique considèrent l'image comme une matrice de pixels indépendants, comme c'est le cas pour l'égalisation d'histogramme ou la correction.

Pour les approches de normalisation qui se basent sur l'extraction de la réflectance, l'inconvénient réside dans la perte d'information liée à la suppression de la composante luminance. Comme nous l'avons déjà cité, l'estimation de la luminance est faite sur une modélisation assez simpliste de l'image. La suppression de cette composante entraîne inévitablement une suppression d'une information importante pour la discrimination des visages. Nous pouvons même remarquer qu'en ne se basant que sur la réflectance du visage, l'information sur la texture est éliminée (Figure 5.8, (x.d) et (x.e)).

Fort de ces remarques, nous proposons, dans la suite du chapitre, une nouvelle méthode de normalisation qui tire avantage de ces deux types d'approches, à savoir

que la correction de la dynamique ne se fera que sur la luminance de l'image².

La méthode que nous proposons est appliquée en plusieurs étapes :

- séparation de la luminance et de la réflectance,
- correction de la luminance par symétrie et transformation non linéaire,
- reconstruction de l'image à partir de la luminance corrigée et de la réflectance originale.

Les formulations de ces étapes sont décrites ci-dessous :

$$I_o = L_o * R_o \quad (5.16)$$

$$L_s = T(L_o) \quad (5.17)$$

$$I_s = L_s * R_o \quad (5.18)$$

avec I_o l'image originale, L_o la luminance, R_o la réflectance et T la transformation de la luminance.

Afin d'effectuer la séparation de la luminance et de la réflectance du visage, nous nous basons sur la méthode du lissage anisotropique [26].

Pour comparer deux images de visage, l'idéal serait d'avoir des images acquises dans exactement les mêmes conditions d'illumination. En général, dans les applications réelles, cette condition ne peut être satisfaite et les conditions d'acquisition, et surtout celle de l'illumination, sont généralement aléatoires.

Plusieurs études ont proposé des méthodes d'alignement de l'illumination entre l'image référence et l'image test [7,23,24,97]. L'inconvénient de ces méthodes est la nécessité d'utiliser plusieurs images sous plusieurs conditions d'illumination pour créer un modèle de la personne.

La méthode que nous proposons a pour but d'aligner l'illumination des images des visages suivant une direction unique. Par ailleurs, l'objet "visage" présente une symétrie axiale verticale ainsi qu'une forme qui peut être approximée à une forme cylindrique ou sphérique. Cet objet peut donc être approximé à un objet à surface lambertienne.

En partant d'une direction d'illumination quelconque (Figure 5.9, (a)), notre but est de transformer l'illumination de l'image originale afin d'avoir une direction d'illumination de référence, comme montré par la figure 5.9, (b).

Le vecteur d'illumination \vec{L} peut être décomposé en somme de deux vecteurs :

$$\vec{L} = \alpha(\vec{L}_a + \vec{L}'_a) \quad (5.19)$$

2. La réflectance est une information intrinsèque du visage due à la nature de la peau, calculée sur chaque pixel de l'image. Une modification de l'information luminance entraînera une modification importante sur le visage.

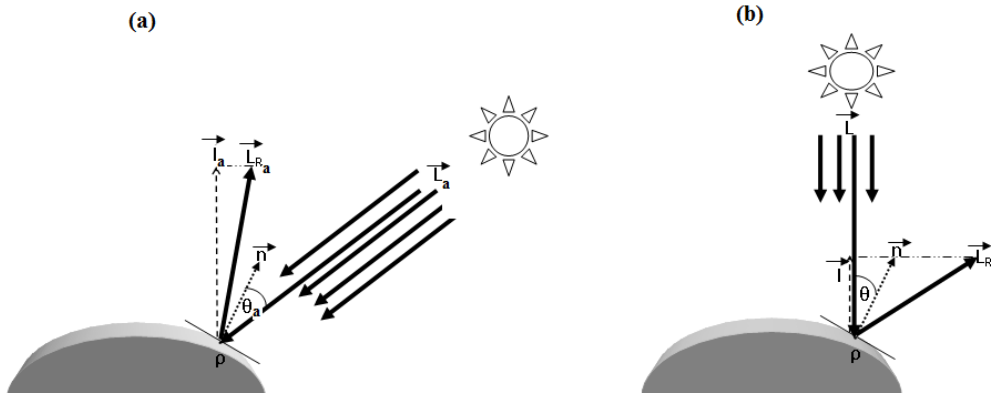


FIGURE 5.9: Schéma des modèles de l'illumination : (a) modèle de l'illumination originale, (b) modèle d'illumination après alignement

avec \vec{L}_a le vecteur d'illumination originale, $\|\vec{L}'_a\| = \|\vec{L}_a\|$ et α un coefficient de gain (Figure 5.10).

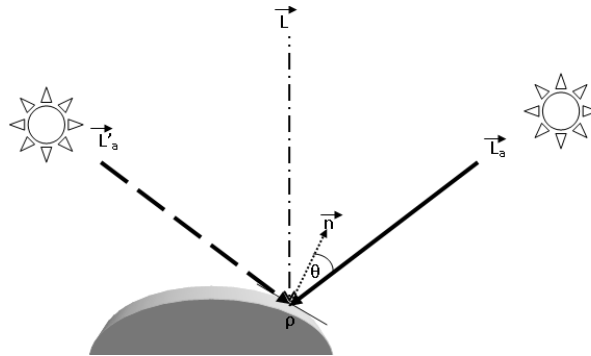


FIGURE 5.10: Décomposition de l'illumination référence en somme de deux illuminations verticalement symétriques.

Soit I l'image du visage avec une illumination verticale de référence L . D'après le modèle (Eq. (5.5)), $I = L * R$ et (Eq. (5.19))

$$I = R * \alpha \sqrt{L_a^2 + L_a'^2} \quad (5.20)$$

La composante L'_a peut facilement être obtenue en considérant l'image symétrique de l'image du visage. En effet, si on considère le visage comme un objet parfaitement symétrique à partir de l'information luminance sur la moitié du vi-

sage l'image miroir, nous permet d'obtenir l'illumination des pixels de l'autre moitié, comme présentée dans l'exemple de la Figure 5.11.

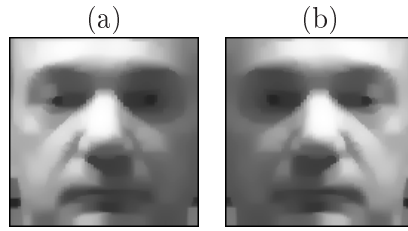


FIGURE 5.11: Exemple d'une illumination originale en (a) et d'une illumination symétrique en (b)

Nous considérons à la fin de cette étape, une transformation logarithmique de l'intensité de l'illumination, ce qui revient à appliquer une correction gamma sur l'image de l'illumination.

Notre nouvelle approche est basée, sur l'hypothèse que le visage est un objet à surfaces parfaitement lambertiennes, sauf que cette caractéristique n'est pas vraie et le visage peut comporter des zones et des surfaces non lambertiennes qui induisent donc inévitablement des ombres (portées ou propres³). Même si notre approche ne prend pas en compte les parties non lambertiennes, elle permet de corriger de manière importante l'illumination de l'image du visage, comme le montrent les exemples de la figure 5.12.

La figure 5.13 montre deux exemples de toute la chaîne de correction de l'image du visage.

3. Une ombre portée est l'ombre d'un objet sur une autre surface et l'ombre propre est l'ombre qui apparaît sur le même objet lorsque un volume de celui-ci se soustrait aux rayons incidents.

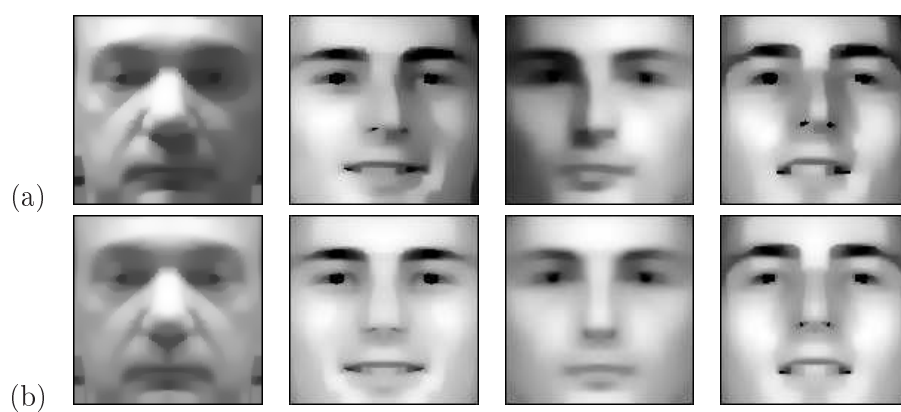


FIGURE 5.12: Exemples de correction de l'illumination par symétrie axiale : (a) l'illumination originale et (b) l'illumination après correction

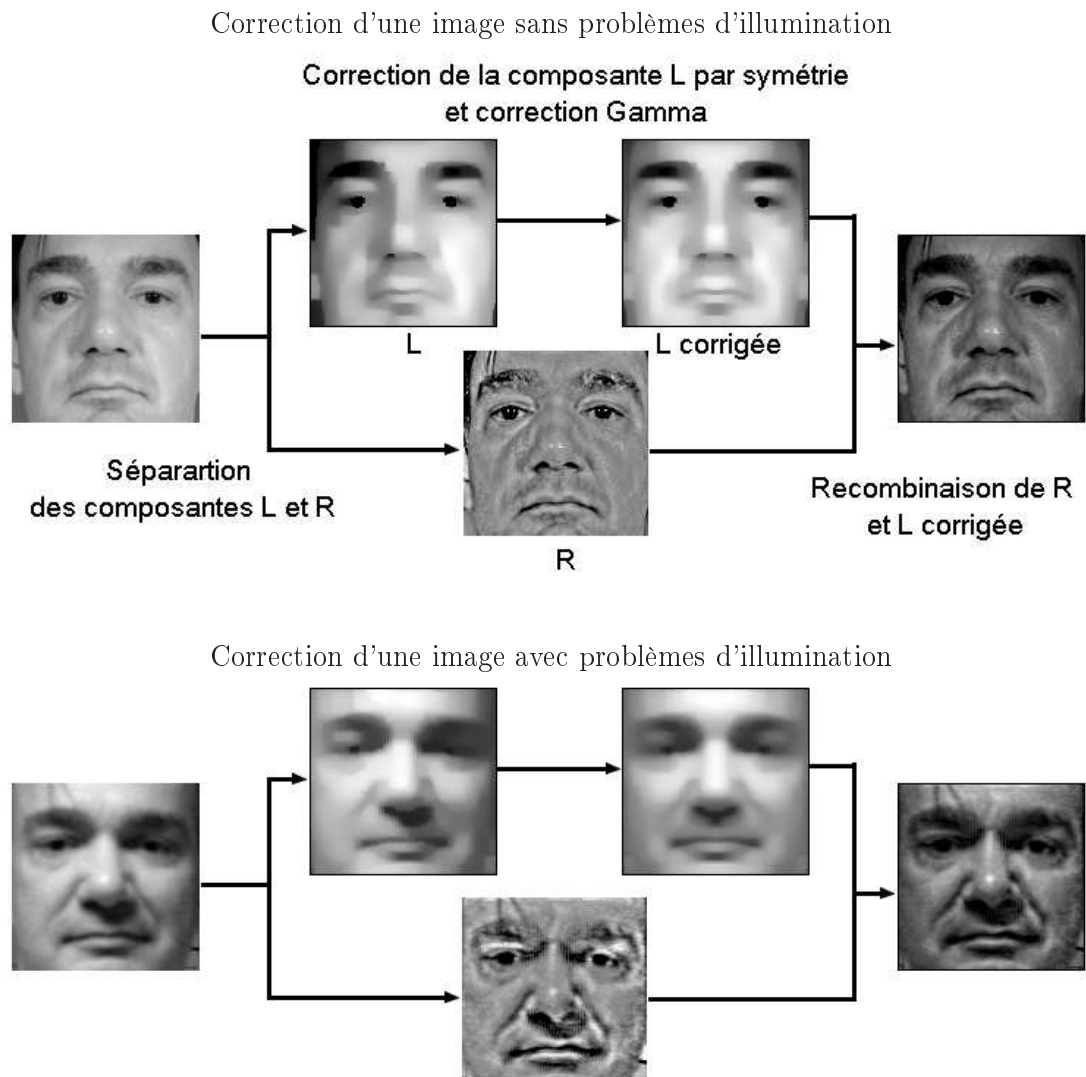


FIGURE 5.13: Correction de l'image du visage par transformation de l'illumination

5.5 Résultats et analyses

Afin de comparer les différentes approches, nous avons considéré les deux bases de données publiques FRGCv2 et BANCA. Pour la base FRGC, nous avons utilisé les deux protocoles de reconnaissance, Expérience 1 et Expérience 4, et nous avons validé les résultats sur la base BANCA en considérant les deux protocoles P et Mc (voir le chapitre 3 section 3.4 pour plus de détails).

5.5.1 Comparaison entre les différentes transformations en niveaux de gris

Dans la première partie du chapitre, nous avons présenté les différentes méthodes d'extraction des niveaux de gris à partir d'une image couleur. Les tableaux 5.1 et 5.2 montrent les résultats de reconnaissance de l'algorithme DLDA appliqué aux différentes approches. Les résultats sont donnés sur la base FRGC (Expérience 1) ainsi que sur la base BANCA (protocole Mc). Ces protocoles "propres" ont été choisis afin de pouvoir comparer les résultats indépendamment des problèmes d'illumination qui peuvent entacher les analyses.

Point de fonctionnement	RVB	TSL	TSV
EER	10.1	7.5	6.5
VR	55.2	64.5	68.2

TABLE 5.1: Résultats de la DLDA donnés pour les différentes méthodes d'extraction des niveaux de gris (RVB, TSL, TSV) sur la base FRGC aux deux points de fonctionnement EER et VR@0.1% de FAR (Expérience 1)

Point de fonctionnement	RVB	TSL	TSV
EER	15.2	13.2	11.1

TABLE 5.2: Résultats de la DLDA donnés pour les différentes méthodes d'extraction des niveaux de gris sur la base BANCA au point de fonctionnement EER (Protocole Mc)

Les performances obtenues en utilisant la composante "Valeur" du système de couleurs TSV présentent un avantage net par rapport à l'approche par moyenne des composantes RVB et un léger avantage par rapport de la composante luminance du système colorimétrique TSL.

La différence de résultats entre le tableau 4.3 et les tableaux 5.1 et 5.2 s'explique par l'utilisation d'un prétraitement par égalisation d'histogramme pour les images de tests du tableau 4.3.

Dans la suite, nous allons considérer la composante "Valeur" du système de couleurs TSV comme méthode d'extraction des niveaux de gris pour les différents tests et protocoles.

5.5.2 Comparaison des différentes méthodes de normalisation

Les résultats de l'algorithme DLDA appliqué aux différentes méthodes de normalisation (égalisation d'histogramme, correction Gamma, Multiretinex, lissage anisotropique et correction de la luminance par symétrie) sont présentés dans le tableau 5.3 pour les protocoles de test de FRGC en terme de VR@0.1% de FAR et dans les figures 5.14 (a) et (b) en terme de courbes ROC. Le tableau 5.4 présente les résultats de la DLDA appliquée à la base BANCA pour les deux protocoles P et Mc. Les résultats sont donnés en terme de EER.

	EH	Gamma	Multiretinex	L. anisotropique	C.L.S
Expérience 1	86% $[\pm 0.5]$	81% $[\pm 0.4]$	75% $[\pm 0.8]$	77% $[\pm 0.8]$	83% $[\pm 0.5]$
Expérience 4	32% $[\pm 1.1]$	33% $[\pm 1.1]$	36% $[\pm 1.1]$	38% $[\pm 1.0]$	44% $[\pm 0.9]$

EH : Égalisation d'histogramme

C.L.S : Correction de la luminance par symétrie.

TABLE 5.3: Résultats de reconnaissance de la DLDA appliquée à différentes méthodes de pré-traitement. Les résultats sont présentés au point de fonctionnement VR @0.1% de FAR sur la base FRGC pour les expériences 1 et 4.

	EH	Gamma	Multiretinex	L. anisotropique	C.L.S
Protocole Mc	12.1% $[\pm 1.2]$	11.5% $[\pm 1.2]$	10.1% $[\pm 1.1]$	9.3% $[\pm 0.9]$	6.5% $[\pm 0.6]$
Protocole P	16.1% $[\pm 1.5]$	15.5% $[\pm 1.4]$	13.2% $[\pm 1.3]$	11.1% $[\pm 1.2]$	9.8% $[\pm 0.9]$

TABLE 5.4: Résultats de reconnaissance de la DLDA appliqué à différentes méthodes de pré-traitement. Les résultats sont présentés au point de fonctionnement EER sur la base BANCA pour les protocoles Mc et P.

Nous pouvons déduire de ces résultats l'importance du choix du pré-traitement, lorsque les conditions d'acquisition sont non contrôlées. En effet une atténuation de la variation de l'illumination entre les images de référence et les images de test est nécessaire afin d'améliorer les performances de la reconnaissance. Lorsque la variation de l'illumination entre l'image de référence et celle de test n'est pas très grande, comme c'est le cas pour l'expérience 1 de FRGC et le protocole Mc

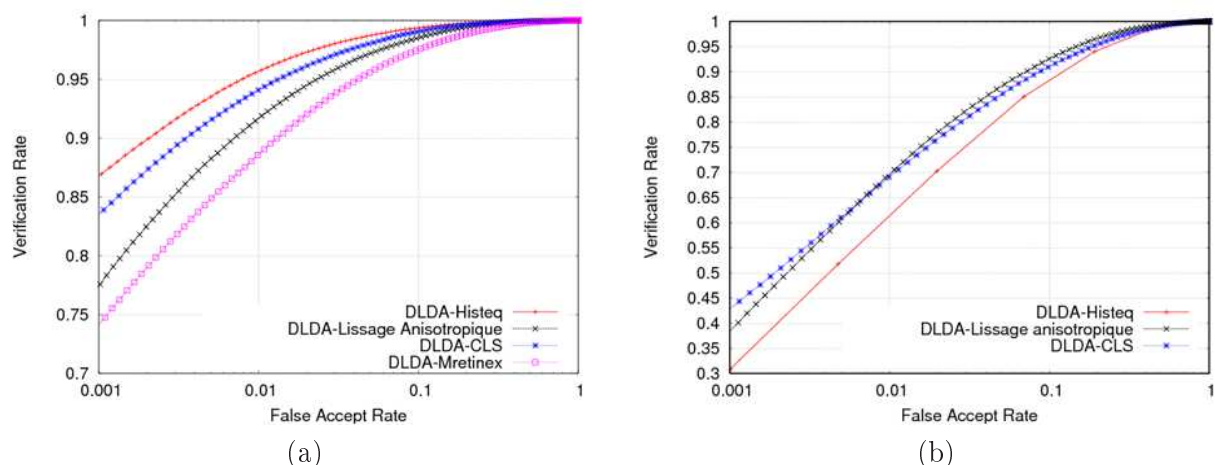


FIGURE 5.14: Courbes ROC de l'expérience 1 (a) et de l'expérience 4 (b) de la base FRGC

de BANCA, une simple égalisation d'histogramme est suffisante. Dans le cas de l'expérience 4 de FRGC et du protocole P de BANCA, les algorithmes de pré-traitement se basant sur la réflectance de l'image ont montré leur robustesse.

La robustesse de la méthode de pré-traitement que nous avons proposée est démontrée par les performances de l'algorithme DLDA.

5.6 Conclusion

Dans ce chapitre nous avons présenté les méthodes de pré-traitement de l'image du visage, qui sont les plus utilisées dans la littérature, à savoir l'égalisation d'histogramme, la correction gamma, le multiscale-retinex et le lissage anisotropique. Nous avons aussi proposé une nouvelle méthode de pré-traitement basée sur la séparation de la réflectance et de la luminance dans un premier temps, puis sur une correction de la luminance par symétrie et par transformation non linéaire et enfin, sur la reconstruction de l'image du visage à partir de la réflectance originale et la luminance corrigée. Cette méthode a montré sa robustesse dans les conditions d'illumination difficiles. En effet, les tests pour les protocoles difficiles de FRGC et de BANCA ont montré que notre approche couplée à la DLDA apporte une réelle amélioration des performances.

Finalement, lorsque les images de référence et de test sont acquises dans des conditions d'illumination et de qualité contrôlées, une simple égalisation d'histogramme permet d'avoir les meilleurs résultats et aucun autre pré-traitement

n'est alors nécessaire. Lorsqu'une variation de l'illumination est présente entre l'image de référence et l'image de test (l'expérience 4 de FRGC et le protocole P de BANCA), un pré-traitement est nécessaire pour limiter l'influence de cette variation.

Chapitre 6

Analyse globale du visage basée sur le filtrage de Gabor

6.1 Introduction

Dans le domaine de la reconnaissance des formes, l'analyse par filtrage de Gabor s'est avérée particulièrement appropriée pour la représentation et la discrimination de la texture. Les caractéristiques directement extraites par filtrage de Gabor, à partir des images des pixels, ont été largement utilisées dans l'identification d'empreintes digitales [45], la segmentation de la texture [34], et particulièrement dans l'identification de l'iris [18]. Comme citées dans le chapitre 2, elles ont été également employées pour l'identification de visage [49, 84, 105].

Dans ce chapitre, nous rappelons tout d'abord la définition des ondelettes de Gabor et nous proposons une étude approfondie des différents paramètres caractérisant ces ondelettes qui permettent d'optimiser les analyses suivant les caractéristiques de l'image du visage à analyser (dimensions, texture...). Dans la deuxième partie, nous introduisons l'utilisation de la phase de Gabor. Nous expliquons le problème principal lié à l'utilisation de cette phase et nous proposons une solution pour remédier à ce problème.

Le système de reconnaissance que nous avons adopté dans notre étude est basé sur l'algorithme de réduction d'espace par DLDA (*Direct Linear Discriminant Analysis*, chapitre 4), appliqué à la représentation d'amplitude et de phase de Gabor. Nous montrons que le couplage de l'amplitude et de la phase corrigée de

la représentation de Gabor apporte une réelle amélioration des performances par rapport à une simple utilisation de l'amplitude.

La dernière partie de ce chapitre est consacrée à l'analyse des résultats des différentes expériences que nous avons effectuées sur les bases FRGC et BANCA et nous donnons les résultats de l'évaluation visage 2D de la campagne IV².

6.2 Filtrage de Gabor et extraction des caractéristiques

La méthode de reconnaissance que nous avons adoptée repose, principalement, sur l'extraction des caractéristiques du visage par filtrage de Gabor 2D. Les filtres de Gabor sont connus comme un moyen d'analyse espace-fréquence très robuste. Cette spécificité a fait des filtres de Gabor un moyen puissant d'analyse de texture et de classification. Les filtres de Gabor analysent la texture d'un objet suivant différentes résolutions et différents angles. Dans le domaine spatial, un filtre de Gabor 2D est une fonction à noyau gaussien modulée par une onde sinusoïdale plane complexe :

$$G(x, y) = \frac{1}{2\pi\sigma\beta} e^{-\pi\left[\frac{(x-x_0)^2}{\sigma^2} + \frac{(y-y_0)^2}{\beta^2}\right]} e^{i[\xi_0 x + \nu_0 y]} \quad (6.1)$$

où (x_0, y_0) est le centre du filtre de Gabor dans le domaine spatial, ξ_0 et ν_0 les fréquences spatiales du filtre, et σ et β les écart-types spatiaux de la gaussienne elliptique le long de x et de y (Figure 6.1). Tous les filtres peuvent être produits à partir d'une ondelette mère par dilatation et par rotation de celle-ci. Chaque filtre a la forme d'une onde plane avec une fréquence f , limitée par un écart-type correspondant à l'enveloppe gaussienne.

Extraction des caractéristiques

La représentation de Gabor d'une image de visage est obtenue par la convolution de l'image avec la famille des filtres de Gabor, définie par $IG_{(r,o)} = I * G_{(r,o)}$ où $IG_{(r,o)}$ est le résultat de la convolution de l'image par le filtre de Gabor à une certaine résolution r et à une orientation o .

La famille de filtres de Gabor est alors caractérisée par un certain nombre de résolutions, d'orientations et de fréquences, qui seront appelées "caractéristiques" par la suite.

Comme on peut le remarquer dans l'équation 6.1, les filtres de Gabor ont une forme complexe qu'il est possible d'exploiter. En effet, étant en quadrature de

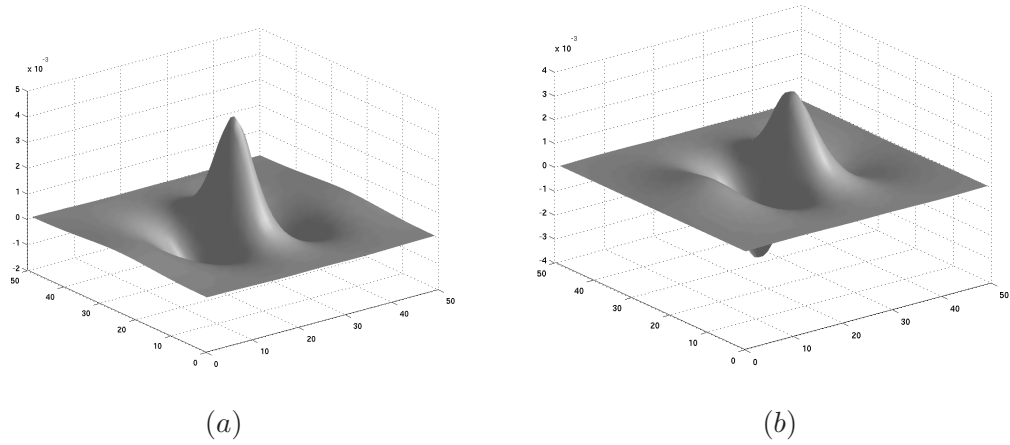


FIGURE 6.1: Partie réelle (a) et imaginaire (b) du filtre de Gabor

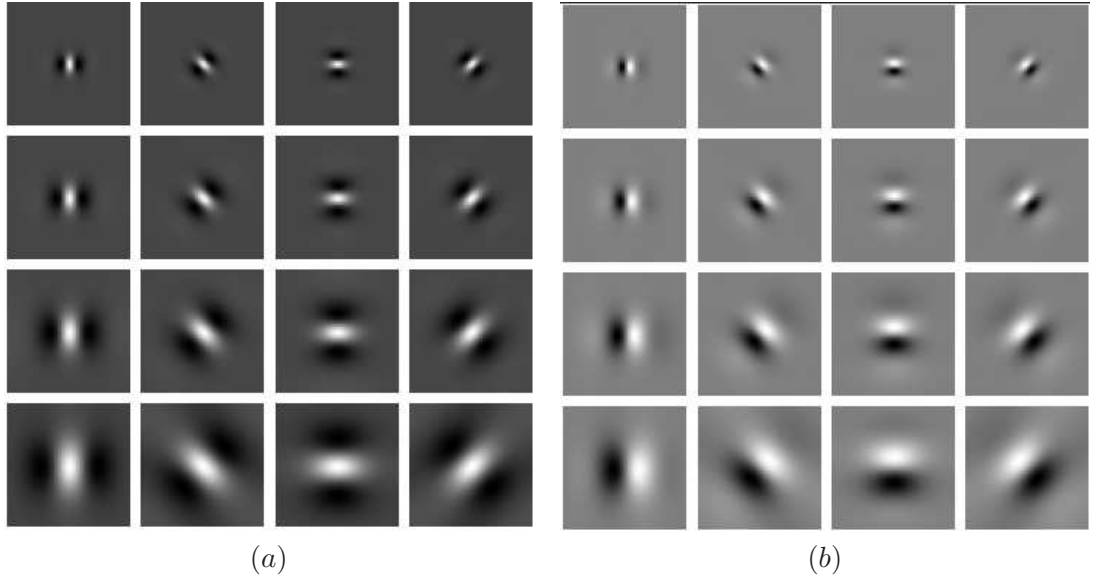


FIGURE 6.2: Parties réelles (a) et imaginaires (b) du filtre de Gabor à 4 niveaux de résolution et selon 4 orientations (les colonnes représentent les orientations et les lignes, les résolutions)

phase, il est important d'utiliser les informations données par la partie réelle et la partie imaginaire des coefficients de Gabor. Deux choix triviaux s'offrent à nous : l'étude de l'amplitude et l'étude de la phase de Gabor. Dans [62], Oppenheim a démontré que la phase des analyses multi-résolutions est plus informative que les amplitudes qui peuvent souffrir des problèmes d'illumination.

L'amplitude et la phase seront notées respectivement par :

$$M(IG_{r,o}(x, y)) = \sqrt{\text{Im}(IG_{r,o}(x, y))^2 + \text{Real}(IG_{r,o}(x, y))^2} \quad (6.2)$$

et

$$P(IG_{r,o}(x, y)) = \arctan\left(\frac{\text{Im}(IG_{r,o}(x, y))}{\text{Real}(IG_{r,o}(x, y))}\right) \quad (6.3)$$

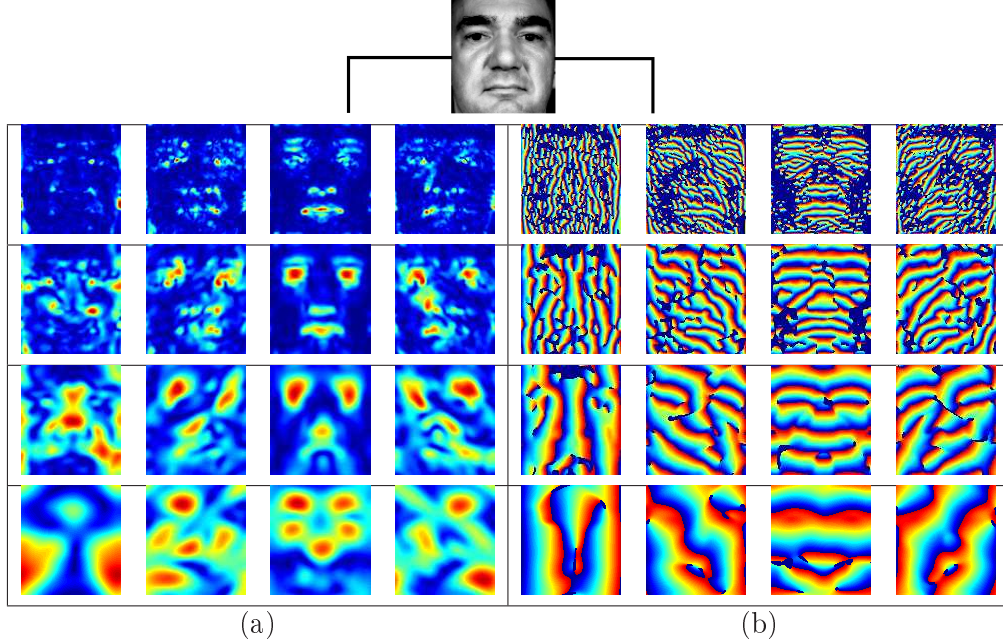


FIGURE 6.3: Résultats de la convolution d'une image de visage avec une famille de 16 filtres de Gabor (4 orientations (horizontales) et 4 résolutions (verticales)). l'ensemble (a) représente les amplitudes et (b) les phases de cette convolution

6.3 Représentation du visage et choix des filtres de Gabor

La majorité des approches se basant sur le filtrage de Gabor pour la reconnaissance de visage (approches globales par réduction d'espace ou approches locales par points caractéristiques) utilisent uniquement l'amplitude de la réponse ou la fusion de l'amplitude et de la partie réelle [49, 84, 105]. L'utilisation de ces représentations est rarement argumentée et la non-utilisation de la phase de Gabor est, dans la plupart des cas, justifiée par une instabilité temporelle de la réponse.

Par exemple, pour l'EGM (*Elastic Graph Matching* [100]) et ses variantes, le calcul de similarité entre deux images de visage se base sur les amplitudes

des réponses des filtres de Gabor. La phase est seulement utilisée pour affiner la détection des noeuds.

Dans [49], l'auteur a publié les meilleurs résultats connus pour la base de données FRGC pour l'expérience 4 avec un taux de VR@0.1% qui s'élève à 78%. Dans son étude, l'auteur utilise la fusion entre amplitude et partie réelle des réponses du filtrage comme représentation du visage, en se basant sur l'approche par réduction d'espace KFA (*Kernel Fisher Analysis*) comme algorithme de reconnaissance.

6.3.1 Influence des caractéristiques de la famille des filtres de Gabor sur les performances de la reconnaissance

Nous commencerons par étudier l'influence des caractéristiques des familles des filtres de Gabor sur les performances de la reconnaissance pour en déduire le choix optimal.

La représentation de l'image considérée, à ce stade du chapitre, est l'amplitude des réponses des filtres de Gabor. L'image d'entrée est une image de visage normalisée géométriquement en fixant une distance de 50 pixels entre les yeux, ce qui équivaut à une taille d'image de 128x128 (pixels). Nous considérons la composante "Valeur" de l'image couleur dans l'espace HSV comme mesure d'intensité du pixel (voir chapitre 5). Une simple égalisation d'histogramme est considérée comme étape de pré-traitement.

L'algorithme de reconnaissance utilisé dans ce chapitre est la DLDA (voir chapitre 4) appliquée à l'amplitude des réponses des filtres de Gabor (figure 6.3(a)). Chacune des 16 images amplitude correspondant à une orientation/échelle est transformée en un vecteur par balayage des colonnes. Ces 16 vecteurs sont ensuite concaténés dans un unique vecteur qui sera considéré comme la nouvelle représentation du visage. La mesure de similarité utilisée est le cosinus entre les représentations projetées sur le sous espace adéquat.

Afin de simplifier l'optimisation du choix des filtres, nous mettons en place un protocole d'évaluation réduit par rapport aux protocoles d'évaluation de FRGC. Pour cela, nous allons considérer un ensemble de 220 personnes avec 10 images par personne (5 images seront utilisées comme référence et 5 images comme test). Pour chaque client, 10 imposteurs, choisis aléatoirement parmi les 219 personnes restantes, lui sont associés. Ce protocole, qui sera noté par la suite PEG (**P**rotocol **E**d'Evaluation de Gabor) comporte 5000 tests intra-classes et 50.000 tests inter-classes.

La base d'apprentissage de l'espace de réduction par DLDA est constituée des données des mêmes 220 personnes. Nous choisissons 5 images par personne, n'appartenant pas à la base de test pour construire les 200 classes d'apprentissage.

Dans cette première phase, nous choisissons de limiter le nombre des niveaux de résolution et d'orientation des filtres de Gabor à respectivement 6 résolutions et 8 orientations.

Le tableau 6.1 présente les différents résultats de reconnaissance en terme d'EER pour chaque niveau de résolution considéré indépendamment des autres niveaux et en variant le nombre d'orientations des filtres par niveau.

Niveau de résolution	2 orientations	4 orientations	8 orientations
1	5.94[±0.45]	2.65[±0.35]	1.41[±0.35]
2	3.46[±0.41]	1.56[±0.36]	1.31[±0.32]
3	3.95[±0.42]	3.05[±0.4]	3.09[±0.4]
4	7.62[±0.54]	6.47[±0.51]	6.68[±0.52]
5	18.56[±1.5]	15.86[±1.1]	13.16[±0.8]
6	28.6[±2.5]	27.8[±2.3]	26.6[±2.3]

TABLE 6.1: Performances en EER [±CC] pour chaque niveau de résolution en fonction du nombre d'orientations choisi par le filtrage de Gabor sur le protocole PEG

Les résultats de cette première expérience montrent clairement que les niveaux de résolution les plus fins (niveau 1 à 3) analysent mieux les caractéristiques discriminantes que les niveaux les plus "grossiers" (niveau 4 à 6). Une augmentation du nombre des orientations ajoute dans la majorité des cas de la robustesse à l'analyse ; pour les niveaux 3 et 4 l'augmentation des orientations reste stable au vue de l'intervalle de confiance . Il est donc utile d'exploiter un nombre maximal d'orientations et de plus l'utilisation des niveaux d'analyse les plus fins (inférieurs à 3 niveaux) favorise la discrimination entre les visages.

L'intérêt de l'utilisation du filtrage de Gabor réside dans la complémentarité des niveaux d'analyse (appelée aussi "analyse multi-résolution"). Le tableau 6.2 présente les résultats de combinaison des différents niveaux en choisissant de fixer à 8 le nombre d'orientation.

Si on considère le point EER comme le point de fonctionnement de notre système, il est clair que les niveaux de résolution supérieurs à 2 n'apportent pas d'amélioration (au vu des résultats dans l'intervalle de confiance). Le comportement du système change radicalement au point de fonctionnement de VR@0.1% de FAR (une sécurité accrue contre l'imposture). Les résultats montrent que deux

Nombre de niveaux	EER [\pm IC]	VR @0.1% de FAR [\pm IC]
1	1.41[\pm 0.35]	80.2[\pm 2.3]
2	0.75[\pm 0.23]	87.2[\pm 2.1]
3	0.76[\pm 0.24]	89.5[\pm 2.1]
4	0.68[\pm 0.23]	97.2[\pm 0.4]
5	0.69[\pm 0.24]	97.3[\pm 0.4]
6	0.71[\pm 0.25]	96.8[\pm 0.4]

TABLE 6.2: Résultats en EER et en VR@0.1% de FAR des combinaisons cumulatives des différents niveaux de résolution avec 8 orientations pour le protocole de test PEG

niveaux de résolution ne sont pas suffisants mais qu'à partir de 4 niveaux de résolution, les résultats se stabilisent. Nous choisissons d'étudier notre système aux points de fonctionnement VR@0.1% de FAR afin de pouvoir comparer nos résultats à ceux de la littérature.

De plus, la stabilité de ces résultats à partir d'un certain niveau de résolution s'explique par la capacité de l'algorithme de la DLDA à extraire les vecteurs de l'espace de réduction en se basant sur les données discriminantes. On remarque, en outre, qu'à partir du niveau 4 de résolution, l'analyse par filtrage de Gabor n'est plus très discriminante (en d'autres termes, il n'y a plus d'information supplémentaire qui permet de séparer les classes).

Pour confirmer cette hypothèse, nous avons mesuré la similarité des axes principaux produits par la DLDA à partir du 4ème niveau de résolution. La figure 6.4 montre les mesures de similarité (cosinus de l'angle) des axes construits par les 4, 5 et 6 premiers niveaux. Le tableau 6.3 donne les moyennes et les écart-types entre les niveaux 4-5 et 4-6.

Niveaux	Moyenne	Écart-type
4-5	0.95	0.01
4-6	0.96	0.01

TABLE 6.3: Moyenne et écart-types des similarités entre les axes principaux de la DLDA construits par les 4, 5 et 6 premiers niveaux

Les résultats de la figure 6.4 (résumés dans le tableau 6.3) qu'à partir d'un certain niveau de résolution (niveau 4), l'information rajoutée n'est plus discriminante. Au vu des résultats, nous avons fixé à 4 le niveau maximal de résolution et à 8 le nombre d'orientations. Ces caractéristiques seront utilisées par la suite.

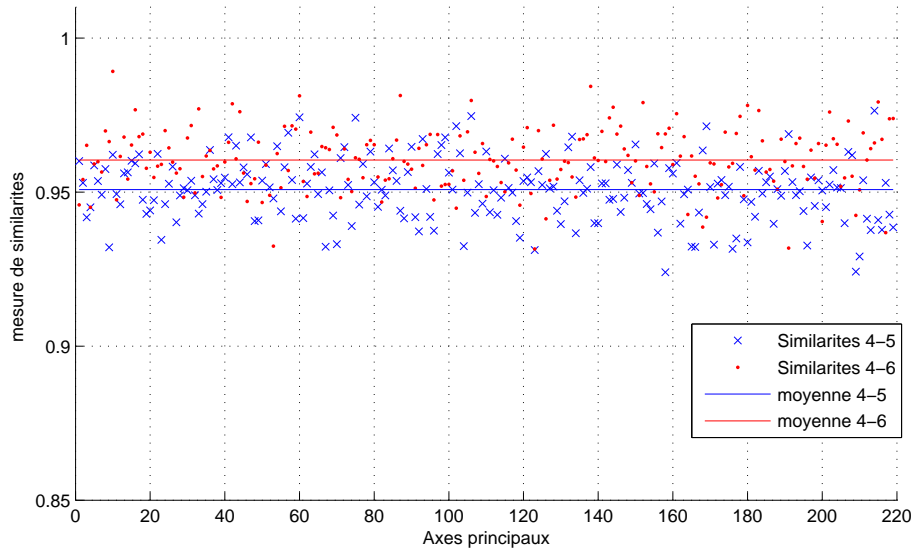


FIGURE 6.4: Mesure de similarités entre les axes principaux de la DLDA construits par les 4,5 et 6 premiers niveaux

6.3.2 Études des performances suivant le choix des représentations de Gabor

Dans la partie précédente, nous avons choisi uniquement l'amplitude des réponses des filtres de Gabor comme représentation du visage. Nous présentons dans cette section les différentes performances obtenues en fonction des représentations de Gabor possibles.

Problème de l'utilisation de la phase de Gabor pour les visages

Lorsque nous considérons une image normalisée du visage (distance fixe entre les centres des yeux), certaines parties du visage n'ont aucune texture informative qui pourrait être analysée par les basses résolutions des filtres de Gabor. Pour ces régions, l'analyse par filtrage de Gabor donne $\text{Real}(\text{IG}_{s,o}) \sim 0$ et $\text{Im}(\text{IG}_{s,o}) \sim 0$. Même si ces valeurs sont très près de 0, l'amplitude de la convolution n'est pas affectée par ce problème, alors que la phase devient une forme indéterminée pour ces régions spécifiques.

Pour éviter les formes indéterminées, nous proposons de sélectionner les régions informatives par seuillage de l'amplitude à chaque point d'analyse.

$$P(IG_{s,o}(x, y)) = \begin{cases} \arctan\left(\frac{\text{Im}(IG_{s,o}(x, y))}{\text{Real}(IG_{s,o}(x, y))}\right) & \text{si } M(IG_{s,o})(x, y) > Th \\ 0 & \text{si } \end{cases} \quad (6.4)$$

où (x, y) sont les coordonnées du point analysé et Th est le seuil de sélection de la phase.

Optimisation du seuil pour la sélection de la phase

Afin d'étudier l'influence du seuillage de la phase en fonction des performances, nous exploitons encore une fois le même protocole d'évaluation noté PEG construit dans la première partie du chapitre. La figure 6.5 montre l'évolution des taux de EER et de VR@0.1% de FAR en fonction du seuil Th .

Les courbes de la figure 6.5 montrent que la variation du taux de reconnaissance en utilisant la phase de Gabor est bien liée aux seuils du filtrage. En ne fixant pas de seuil de filtrage, les résultats de la reconnaissance se dégradent considérablement (figure 6.5), ils décroissent à 79% de VR@0.1% de FAR depuis la valeur maximale de 93%. La même étude a été publiée dans [68] mais avec les tests pour l'expérience 1 et l'expérience 4 de FRGC et le même comportement est observé pour le taux de VR@0.1% de FAR.

Comparaison des performances suivant la représentation du visage

Le tableau 6.4 et la figure 6.6 résument les performances en EER de la DLDA appliquée aux différentes représentations de Gabor (amplitude, phase, partie réelle et partie imaginaire) pour le protocole PEG. Nous fixons les paramètres de la famille des filtres de Gabor à 4 niveaux de résolution et à 8 orientations.

Représentation de Gabor	EER	VR@0.1% de FAR
Amplitude	0.68%	93.4%
Phase	1.76%	89.20%
Réelle	2.12%	80.40%
Imaginaire	2.11%	80.41%

TABLE 6.4: Résultats en EER et en VR@0.1% de FAR des différentes représentations de Gabor

Les résultats montrent clairement que l'utilisation de la partie réelle ou de la partie imaginaire est beaucoup moins discriminante si les deux parties sont

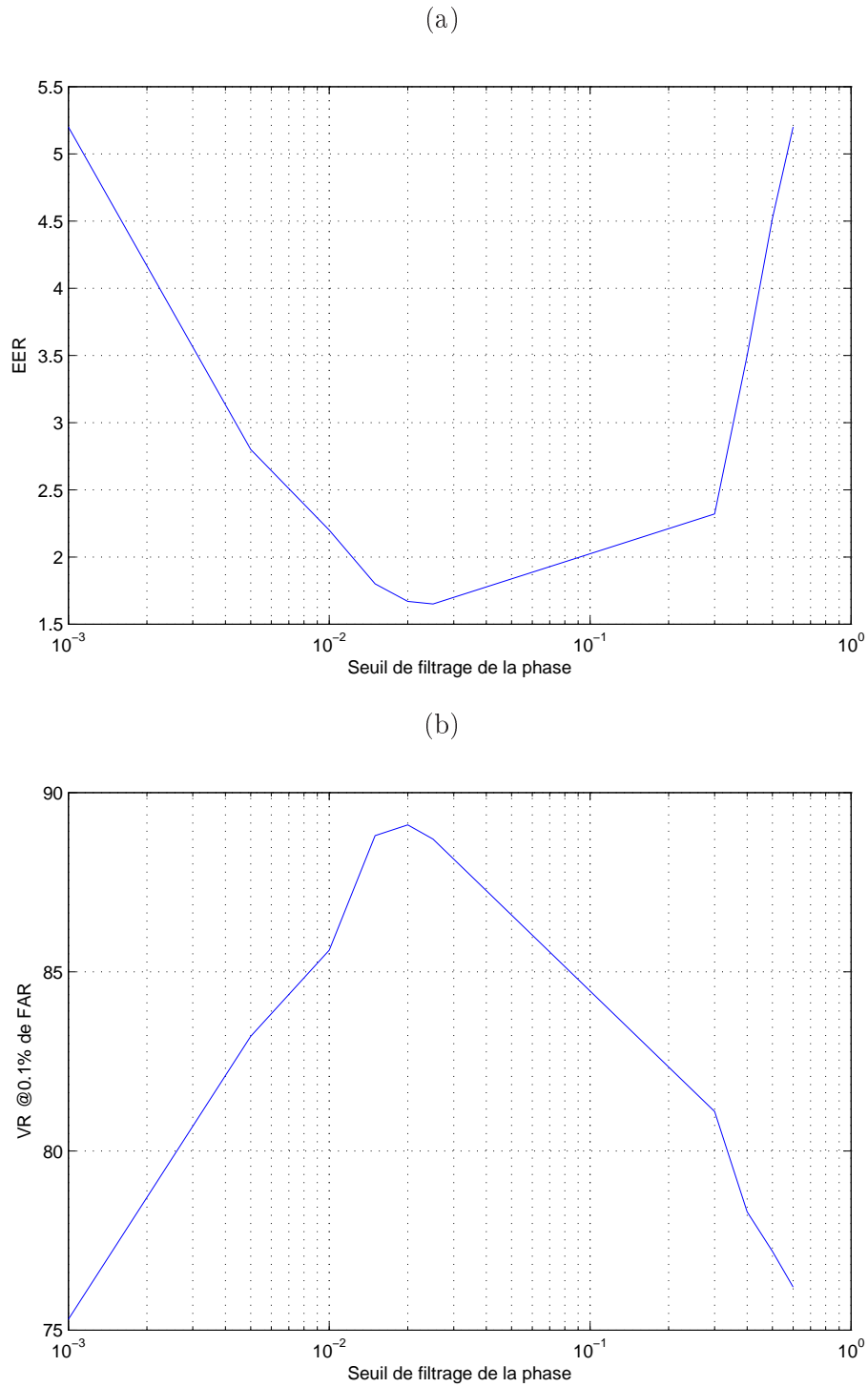


FIGURE 6.5: Évolution des taux de EER (a) et de VR@0.1% de FAR (b) en fonction du seuil Th pour le protocole PEG

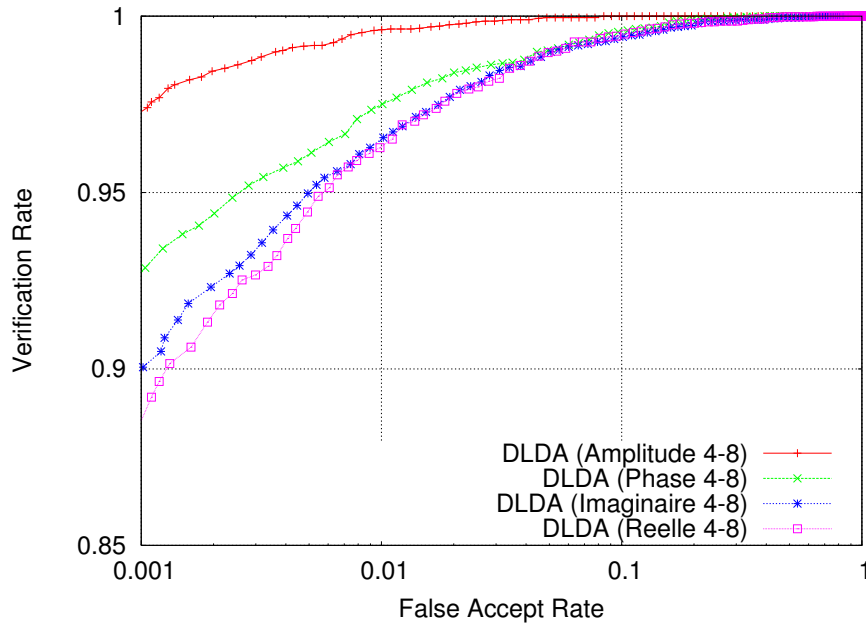


FIGURE 6.6: Courbes ROC des différentes représentations de Gabor pour le protocole PEG avec une famille de filtres à 4 résolutions et 8 orientations

utilisées séparément. La combinaison de la partie réelle et de la partie imaginaire (en amplitude ou en phase) améliore les performances de façon significative.

6.3.3 Fusion de l'amplitude et de la phase des représentations de Gabor

Pour notre approche, nous avons étudié la combinaison de l'amplitude et de la phase, motivée par le fait que l'information utile de la texture est située dans la phase de l'analyse par filtres de Gabor [63] ainsi que par le succès de l'utilisation de la phase en identification de l'iris [18]. La fusion de l'amplitude et de la phase se fait par concaténation dans un seul vecteur de ces deux informations comme expliqué dans la section 6.3.1. Ce vecteur est ensuite réduit par échantillonnage¹ et sera considéré comme la nouvelle représentation du visage.

Le tableau 6.5 présente les différents résultats de fusion obtenus sur le protocole PEG.

Les résultats montrent clairement que la fusion avec la phase de Gabor apporte une réelle amélioration des performances. Une validation de ces résultats avec

1. pour des raisons de limitation de mémoire nous ne gardons qu'un huitième de l'information.

Représentation de Gabor	EER	VR@0.1% de FAR
Amplitude	0.68% $[\pm 0.08\%]$	93.4% $[\pm 0.18\%]$
Amplitude + Phase	0.45% $[\pm 0.02\%]$	96.9% $[\pm 0.11\%]$
Amplitude + Partie Réelle	0.64% $[\pm 0.05\%]$	94.2% $[\pm 0.15\%]$
Amplitude + Partie Imaginaire	0.67% $[\pm 0.08\%]$	93.4% $[\pm 0.18\%]$

TABLE 6.5: Résultats en EER et en VR@0.1% de FAR des différentes fusions des représentations de Gabor

d'autres méthodes de classification est nécessaire pour confirmer l'apport réel de la fusion amplitude-phase.

6.3.4 Validation de la fusion de l'amplitude et de la phase corrigée sur des algorithmes de réduction d'espace à base de noyau

Les algorithmes de reconnaissance des visages par projection dans les espaces réduits ont connu une évolution significative avec l'application des noyaux non linéaires à partir de 2000 [56, 57, 81]. Ces méthodes ont démontré, dans la littérature, leur supériorité par rapport aux approches linéaires classiques. Nous pouvons citer comme méthode la GDA (*General Discriminant Analysis*) [56] et l'analyse non linéaire discriminante de Fisher (KFDA ou KFA, *Kernel Fisher Discriminant Analysis*) [57]. L'idée principale qui soutend ces approches non linéaires est que les représentations des visages sont non linéairement séparables. Une projection par noyaux non linéaires des visages dans un espace de plus grande dimension, devrait donc permettre de mieux les séparer. L'efficacité de cette technique a été confirmée par l'amélioration des résultats dans les études déjà cités ainsi que dans [49]. Dans cette dernière publication Liu en 2006 a utilisé la KFA appliquée aux représentations constituées par l'amplitude et la partie réelle des réponses de Gabor.

NB :La complexité algorithmique et les longs délais d'exécution de ces algorithmes (développés à l'INT et non optimisés), ne nous permettent pas d'exploiter de façon efficace ces approches sur des grandes base de données comme FRGC. Toute fois, nous présentons, dans la suite, les résultats de performance de la GDA, de la KFA ainsi que les deux approches globales classiques, la LDA et la DLDA, appliquées aux paramètres résultants de la fusion de l'amplitude et de la phase corrigée pour le protocole PEG.

Les courbes ROC de la figure 6.7 confirment la supériorité des approches se basant sur la réduction d'espace par noyaux non linéaires par rapport aux

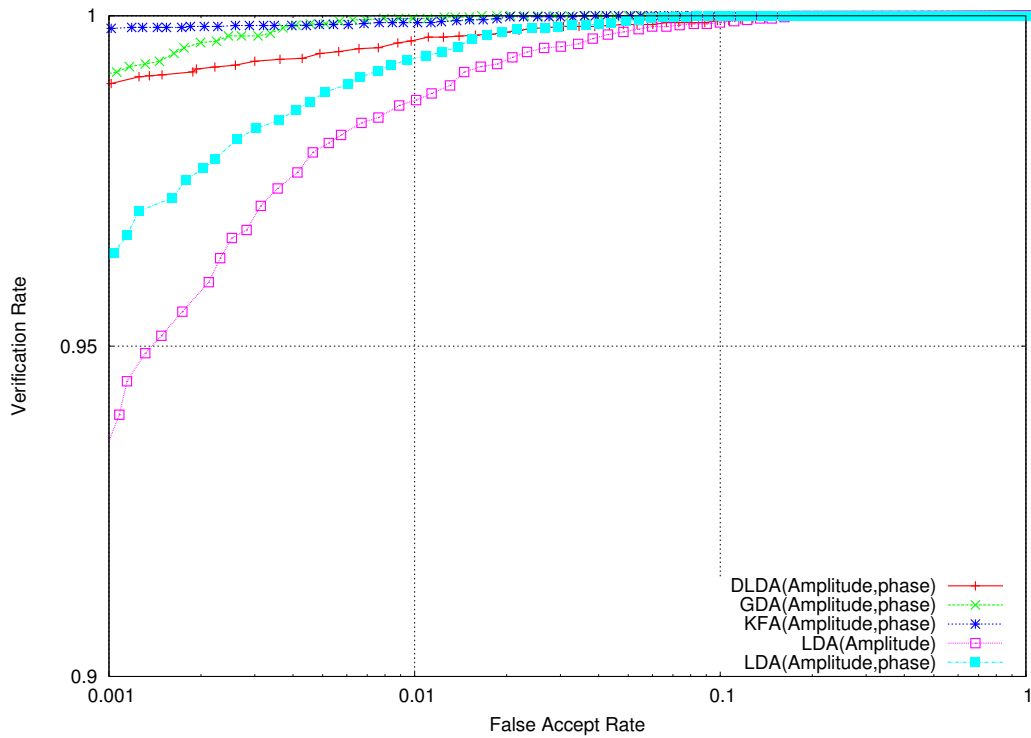


FIGURE 6.7: Courbes ROC des différentes méthodes pour le protocole PEG

approches linéaires. Nous pouvons constater aussi que l'algorithme de la DLDA, appliqué à la fusion de l'amplitude et de la phase, réduit de façon significative l'écart de performance avec les approches non linéaires.

L'adoption de cette approche (DLDA appliquée à la fusion de l'amplitude et de la phase) nous semble un bon compromis entre performance et complexité algorithmique.

6.4 Résultats sur la base FRGC et positionnement de notre système par rapport à la littérature

La base FRGC est la base publique la plus importante en terme de données disponibles. Nous avons appliqué notre approche (fusion amplitude et phase de Gabor) sur les deux expériences 1 et 4 afin de pouvoir comparer correctement nos résultats par rapport à ceux de la littérature.

Pour ces expériences, nous avons choisi d'utiliser la normalisation photométrique par lissage anisotropique (voir chapitre 5 ou [26]) car la plupart des algo-

rythmes de la littérature présentés utilisent cette même normalisation.

Algorithme	Expérience	VR@0.1% de FAR [\pm IC]
Résultats de la fusion (amplitude et phase)		
LDA	Exp1	87.62% [± 0.14]
	Exp4	50.22% [± 0.31]
DLDA	Exp1	92.62% [± 0.1]
	Exp4	60.25% [± 0.31]
GDA	Exp1	94.9% [± 0.08]
	Exp4	72.2% [± 0.21]
KFA	Exp1	95.3% [± 0.08]
	Exp4	74.2% [± 0.20]
Résultats de la littérature sur la base FRGC		
KFA [49]	Exp1	–
	Exp4	78.2%
HEGM (Cognitec)	Exp1	–
	Exp4	60.2%
Fusion hiérarchique de classifieurs [88]	Exp1	–
	Exp4	86%

TABLE 6.6: Résultats de la reconnaissance par fusion amplitude-phase en taux de VR@0.1% de FAR pour les deux expériences 1 et 4 de la base FRGC ainsi que quelques résultats de la littérature

Les résultats du tableau 6.6 confirment les résultats obtenus dans la section précédente. Les approches à base de noyaux apportent une réelle amélioration des performances de reconnaissance des visages dans les deux conditions de tests. Exemple pour un gain relatif de performance de $\sim 38\%$ entre la LDA et la KFA pour l'expérience 1 et de 35% pour l'expérience 4.

Nous constatons aussi une différence de résultats entre notre approche par KFA appliquée à la fusion de l'amplitude et de la phase et les résultats de l'approche de Liu [49] qui se base sur l'amplitude et la partie réelle (74.2% de VR@0.1% de FAR contre 78.2%). Ceci pourrait s'expliquer par une différence d'implémentation de l'algorithme ou dans les paramètres de la normalisation photométrique utilisée.

6.5 Résultats sur d'autres bases

Nous présentons dans cette section les résultats sur la base BANCA ainsi que sur la base IV².

6.5.1 Résultats sur la base BANCA

Le tableau 6.7 résume les résultats de la DLDA appliquée à la fusion amplitude-phase de réponses de la famille de 32 filtres de Gabor de 4 résolutions et de 8 orientations pour le protocole P.

	WER(0.1)		WER(1)		WER(10)		Av. WER%
	G1	G2	G1	G2	G1	G2	
DLDA fusion	4.3	4.5	10.30	11.75	3.80	4.53	6.53
Résultats de la littérature							
Univ Nottingham	1.55	1.77	6.67	7.11	1.32	1.58	3.33
Univ Tsinghua	1.13	0.73	2.61	1.85	1.17	0.84	1.39

TABLE 6.7: Résultat en WER (*Weighted Error Rate*) de la DLDA appliquée à la fusion amplitude-phase de réponses de la famille de 32 filtres de Gabor sur la base BANCA (protocole P)

D'après le tableau 6.7, les deux seuls systèmes qui surpassent notre approche sont les systèmes de l'université de Nottingham et celui de l'université de Tsinghua.

6.5.2 Résultats sur la base IV²

Comme évoqué dans le chapitre 3, nous présentons ici les résultats comparatifs de la première évaluation IV². Nous rappelons, dans le tableau 6.8, les caractéristiques des différents protocoles d'évaluation définis et nous donnons, dans le tableau 6.9, les résultats publiés dans [67].

Expérience	2D-Exp1	2D-Exp2	2D-Exp3	2D-Exp4
Sessions	Mono	Mono	Mono	Multi
Qualité	Haute	Haute	Basse	Haute
Variation d'expression	Faible	Non	Non	Non
Variation d'illumination	Non	Oui	Non	Non
N. Intra-classes	2595	2502	1654	1796
N. Inter-classes	2454	2362	1598	1796

TABLE 6.8: Protocoles d'évaluation pour la base IV²

Les résultats de cette première évaluation indépendante (Tableau 6.9) confirment les résultats que nous avons obtenus tout au long de ce chapitre, à savoir que l'utilisation des représentations de Gabor apporte une réelle amélioration des résultats

Participants	2D-Exp1	2D-Exp2	2D-Exp3	2D-Exp4
Participant-1(PCA)	6,6	20,7	26,5	20,2
BioSecure(PCA)	7,3	21,6	17,8	13,6
Participant-2(Mod PCA)	5,3	20,9	27,0	19,4
Participant-3(LDA)	3,7	22,5	27,3	21,7
Résultats de notre approche				
(LDA Amplitude de Gabor)	4,4	15,1	15,9	10,2
(LDA Amplitude et Phase de Gabor)	4,2	12,0	12,5	8,3
(DLDA Amplitude et Phase de Gabor)	3,2	10,1	10,2	6,4

TABLE 6.9: Taux de EER pour les différents algorithmes sur la base IV² (visage 2D) participant à la première évaluation

et ceci, pour les diverses conditions de tests. La fusion de l'amplitude et de la phase corrigée de Gabor améliore sensiblement les résultats par rapport à ceux obtenus par l'utilisation de l'amplitude seule.

A noter ici que pour l'évaluation nous avons participé avec l'algorithme LDA. Les résultats sur la DLDA appliquée aux représentations de Gabor ont été obtenus après l'évaluation officielle.

6.6 Conclusion

Dans ce chapitre, nous avons proposé d'utiliser l'analyse espace-fréquence du visage à l'aide des ondelettes de Gabor. Cette information est extraite de l'image des pixels par application d'une famille de filtres à différentes résolutions et à diverses orientations.

Dans la majorité des études de la littérature, seule l'amplitude de la réponse des filtres de Gabor a été employée alors que la phase est omise. Dans ce chapitre, nous avons aussi expliqué les raisons de la limitation de l'utilisation de cette phase et nous avons fourni une solution simple pour pallier cette limitation. Nous nous sommes également intéressés aux paramètres des différents filtres de Gabor comme les orientations, les résolutions et les fréquences. Nous avons observé que l'algorithme de la DLDA est capable d'extraire l'information discriminante de cette nouvelle représentation même si une redondance de l'analyse spatio-fréquentielle apparaît à partir d'une certaine résolution d'analyse, ce qui permet de limiter naturellement le nombre de niveaux à utiliser lors du traitement.

Nous avons également proposé d'utiliser une nouvelle représentation du visage basée sur la fusion de l'amplitude et de la phase corrigée. Les résultats des diffé-

rentes évaluations sur les deux bases de données FRGC et BANCA ont confirmé la robustesse de cette fusion. Dans la première campagne d'évaluation de la base IV², cette méthode a montré sa supériorité par rapport à plusieurs algorithmes classiques comme la LDA, la PCA et la PCA modulaire. L'amélioration est surtout remarquable lorsque les conditions des tests sont dégradées.

Les taux de performance de l'approche que nous avons adoptée sont bien meilleurs lorsque les images de référence et de test sont de bonne qualité, que lorsqu'elles sont acquises dans des conditions dégradées. Nous étudierons dans le chapitre suivant de nouvelles stratégies de reconnaissance adaptées à la qualité des images, tout en utilisant la même approche que celle étudiée dans ce chapitre.

Chapitre 7

Mesures de qualité pour le visage

Qu'entendons-nous par qualité ? En biométrie, deux définitions ressortent de la littérature générale : pour la première définition, un échantillon biométrique est considéré de bonne qualité s'il convient à un test de reconnaissance. Ce point de vue peut ne pas être en accord avec une conception humaine de la qualité. L'exemple le plus représentatif reste celui de l'empreinte digitale. Si, par exemple, un observateur voit une image d'empreinte avec des crêtes claires, un faible niveau de bruit et un bon contraste, il peut raisonnablement dire que l'échantillon est de bonne qualité. Toutefois, si cette image ne contient pas assez de minuties, cet échantillon peut ne pas être exploitable par la plupart des systèmes de reconnaissance basés sur les minuties.

La deuxième définition se rapporte plus à la qualité de la perception de l'échantillon biométrique à la sortie des capteurs, comme par exemple, les images de visages sont classées de bonne ou de mauvaise qualité suivant la netteté, le niveau du contraste et l'équilibre de l'illumination. Plusieurs études ont présenté des approches pour la mesure de la qualité d'un visage dans une image. La majorité des critères étudiés dans la littérature traitent deux aspects de la qualité : l'aspect géométrique et l'apparence. Le premier aspect regroupe des critères comme la pose, la distance entre les yeux, la présence ou l'absence de lunettes. Le deuxième aspect traite des problèmes de contraste, de netteté et d'illumination.

Dans [22], par exemple, Gao a proposé une méthode de mesure de qualité basée

sur 2 niveaux. Le premier niveau concerne la symétrie du visage et le deuxième concerne la symétrie de l'illumination en se basant sur les histogrammes LBP (*Local Binary Pattern*). Son étude montre une bonne classification des différentes situations par cette approche.

L'étude publiée par Krzysztof [43] propose une approche pour estimer la confiance du score de vérification en se basant sur la qualité de l'échantillon. Pour cela, il propose une mesure de qualité standard pour la netteté de l'image et une nouvelle mesure de qualité basée sur la corrélation entre une image du visage test et celle du visage moyen calculé à partir d'une base d'apprentissage, dont la qualité est considérée comme référence.

Dans [72], l'auteur présente une approche pour intégrer les mesures de qualité du visage dans la fusion des scores de plusieurs systèmes de reconnaissance par le visage. Il introduit pour cela une composante appelée "état de qualité" qui est un groupement des échantillons par classe de qualité. Il utilise cette information pour la normalisation des scores avant leur fusion. Cette approche semble améliorer les résultats par rapport aux fusions classiques utilisées.

Même si les études académiques sur les mesures de qualité, spécifiques aux images du visage, ne sont pas très nombreuses, la mesure de la qualité des échantillons biométriques est considérée, en général, comme une composante à part entière des systèmes biométriques. Des normes internationales ont vu le jour afin de définir certains critères de qualité. Pour le moment, ces normes comme les ISO/IEC WD 19794-[4,5,6,11,13 et 14] qui concernent respectivement les empreintes, le visage, l'iris, la signature, la voix et l'ADN, proposent une normalisation des formats d'échanges de données et donnent des recommandations très strictes pour l'acquisition de chaque type d'échantillon biométrique. Pour la norme ISO/IEC WD 19794-5 spécifique aux visages, les recommandations concernent les rapports géométriques du visage dans l'image, la distance entre les yeux, la pose, l'utilisation d'un arrière plan uniforme, la mise au point de la caméra ainsi que la calibration des couleurs. Dans cette norme, une étude sur l'influence de la pose (rotation planaire ou spatiale) ainsi que l'influence de la distance minimale entre les deux yeux sont présentées et cette norme recommande une rotation planaire maximale de $\pm 5^\circ$, une rotation spatiale maximale de $\pm 8^\circ$ ainsi qu'une distance minimale de 90 pixels entre les centres des deux yeux. Pour la suite du chapitre, on notera "image passeport", une image qui possède ces caractéristiques.

De nouvelles normes ISO spécifiques à la quantification de la qualité des échan-

tillons biométriques sont en cours de développement. Dans les rapports techniques pour l'élaboration de la norme ISO/IEC CD 29794-5 spécifique aux images de visages, les auteurs recommandent la quantification de différentes caractéristiques de l'image du visage :

- la mesure de la symétrie de l'illumination du visage : cette symétrie pourra être calculée sur les niveaux des pixels des images, ou sur les images après filtrage par des filtres de Gabor ou par LBP (*Local Binary Pattern*) ou aussi des filtres ordinaux.
- la mesure de la symétrie de la pose du visage.
- la luminosité de l'image : en étudiant différentes statistiques de l'histogramme (moyenne, variance, asymétrie, coefficient d'aplatissement (coefficient Kurtosis)..).
- le contraste de l'image : les auteurs proposent d'utiliser le coefficient de Weber, le coefficient de Mickelson ou de travailler sur l'énergie de l'image par transformation de Fourier.
- la netteté de l'image par calcul du gradient.

Les auteurs du rapport préconisent la quantification de certaines caractéristiques du visage relatives à l'apparence comme l'expression, le maquillage, le port des lunettes, sans proposer des recommandations pour cette quantification.

La stratégie pour la reconnaissance du visage dans la littérature est basée sur le développement d'algorithmes qui sont supposés être robustes face aux sources de bruit et de dégradation. Nous proposons, dans cette étude, une stratégie qui permet d'adapter les systèmes de reconnaissance en fonction de la qualité des images rencontrées. De cette façon, nous espérons un gain en performance, évidemment aux prix d'une complexité un peu plus importante du système complet.

Ce chapitre sera organisé de la façon suivante : nous présenterons différentes mesures de qualité pour caractériser certaines dégradations possibles de l'image du visage (comme le flou et l'illumination). Ces mesures seront ensuite fusionnées pour aboutir à une mesure de qualité unique du visage. Les trois mesures sont le déséquilibre de l'illumination dans une image de visage, la caractérisation de l'histogramme de la luminance d'un visage par SVM et la mesure de la netteté par approche fréquentielle locale pondérée. Ces nouvelles méthodes de quantification de la qualité, que nous proposons, seront comparées à des méthodes classiques déjà utilisées. Le critère de la pose n'est pas pris en compte dans cette étude car

nous nous intéressons essentiellement à l'étude de la reconnaissance de l'image de visage de face. Pour la variation d'expression, nous montrons que dans le cas d'une légère expression, les performances de la reconnaissance restent stables.

Dans la deuxième partie de ce chapitre, nous montrons l'influence de la qualité de l'image du visage sur les performances en reconnaissance et nous présentons une stratégie de reconnaissance basée sur la qualité. Dans cette partie, nous analyserons l'amélioration des résultats grâce à cette approche.

Dans la dernière partie de ce chapitre, nous présentons les résultats de reconnaissance du visage dans le cas où l'image de référence et l'image de test sont de type dégradé et nous analyserons l'apport de l'utilisation des différentes mesures de qualité dans ce cas difficile. Le cas d'une évaluation avec plusieurs images de référence et de test par personne est aussi étudié dans cette dernière partie.

Tout au long du chapitre, nous validerons nos différents résultats sur la partie évaluation de la base FRGCv2 et nous étudierons aussi la pertinence de nos mesures ainsi que notre stratégie de reconnaissance sur la base BANCA.

7.1 Mesures de qualité

Pour étudier les différentes mesures de qualité, nous avons choisi la partie développement de la base de visage FRGCv2. Cette partie de la base est constituée de 12760 images au total. Deux sous-bases de visage, de type différent, seront utilisées pour l'étude des différentes mesures de qualité [Figure 7.14]. La première base est constituée de 4600 images contrôlées [[BQC] ("Base à Qualité Contrôlée")] de 220 personnes et la deuxième est constituée de 4600 images non contrôlées (illumination et netteté diverses) de 220 personnes [[BQN] ("Base à Qualité Non Contrôlée"]].

Les bases [BQC] et [BQN] contiennent chacune 50% d'images dans lesquelles le visage présente une expression neutre et 50% avec un léger sourire. L'algorithme de reconnaissance de visage DLDA appliqué aux analyses par filtres de Gabor (voir chapitre 4) est utilisé pour cette étude.

7.1.1 Mesure du déséquilibre de l'illumination par distributions locales

La première mesure de qualité que nous proposons est basée sur la mesure des distributions de l'illumination sur le visage. L'image considérée est une image de

visage normalisée géométriquement avec centrage de l'histogramme.

Un visage avec une illumination homogène aura une distribution uniforme des niveaux d'intensité des pixels de la peau. Les mesures statistiques des niveaux d'intensité par zone donnent une indication sur les distributions de l'illumination sur la totalité du visage. Les mesures statistiques étudiées sont la moyenne des niveaux de gris et la médiane. Soit Z_i les zones du visage [Figure 7.1]. La mesure de qualité $Q1$ est définie par $Q1 = std(\mu_i)$ avec μ_i , la mesure statistique sur les intensités de la zone Z_i . Par la suite, on notera $Q1_a$ la mesure de qualité se basant sur les distributions des moyennes et $Q1_b$ la mesure de qualité se basant sur les distributions des médianes.

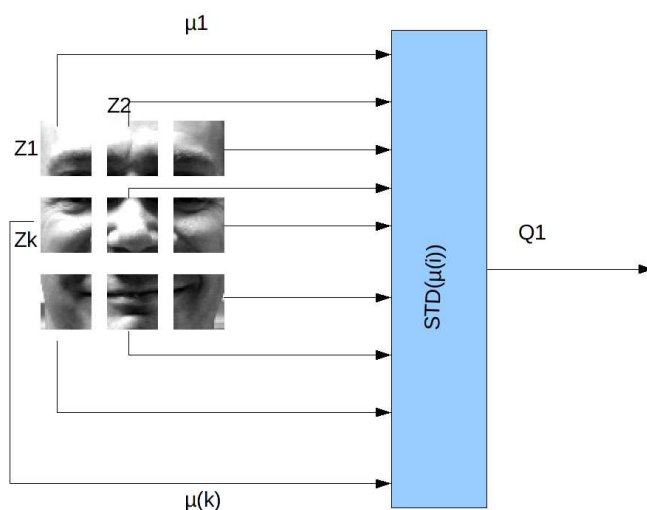


FIGURE 7.1: Principe de calcul d'une mesure de qualité globale ($Q1_a$ ou $Q1_b$) à partir des mesures locales (moyennes ou médianes, respectivement)

La mesure de la moyenne des niveaux locaux, bien qu'elle semble la plus intuitive, pose un certain nombre de problèmes, à savoir que les zones Z_i sélectionnées ne contiennent pas que de l'information sur la texture du visage. Elles contiennent entre autres des parties du visage, comme la bouche, les yeux, les sourcils, qui peuvent induire une erreur sur la caractérisation du niveau de l'illumination de la peau.

L'aspect 3D du visage implique aussi une introduction de bruit, comme les ombres projetées des parties qui constituent le visage. Les résultats de la figure 7.2 montrent les distributions de la mesure de qualité $Q1_a$ des deux bases d'images [BQC] et [BQN].

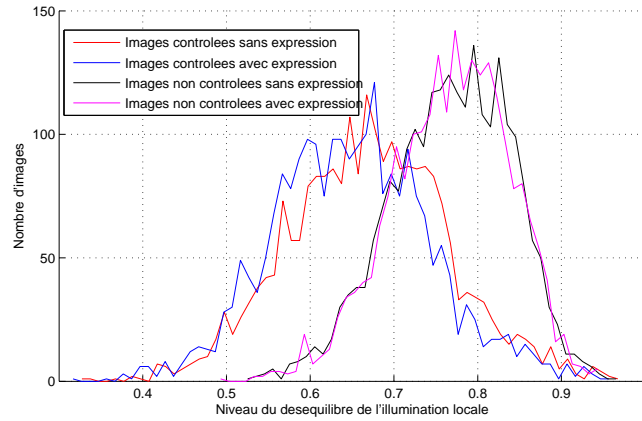


FIGURE 7.2: Distributions de l'illumination par mesure des moyennes locales, $Q1_a$, pour les bases [BQC] et [BQN]

La deuxième mesure statistique considérée est la médiane locale. Cette mesure est plus proche d'une quantification réelle de la texture de la peau si on considère que dans chaque zone Z_i , la peau couvre plus de 50% de la zone. L'utilisation de la médiane donne une indication plus précise sur le niveau de l'illumination de la zone.

Les résultats de la figure 7.3 montrent les distributions de la mesure de qualité par le déséquilibre des médianes locales, $Q1_b$, pour les deux bases d'images [BQC] et [BQN].

Le point faible de cette approche est qu'elle quantifie les distributions des niveaux d'intensité, mais cette quantification peut, dans certains cas, ne pas être très efficace. Une image saturée et une image complètement sombre en sont des exemples. Cette image aura une distribution homogène des illuminations locales et dans ce cas, une erreur de classification peut se produire, puisqu'une telle image ne peut être considérée comme étant de bonne qualité.

Dans la suite nous choisissons la mesure de la variance des médianes locales, comme mesure du déséquilibre de l'illumination.

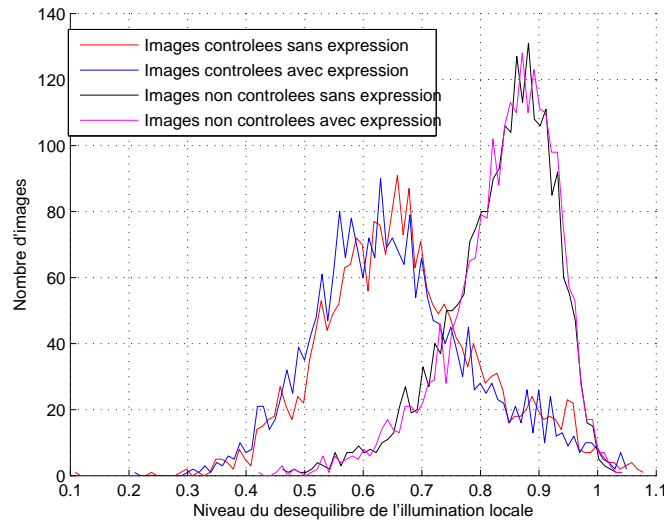


FIGURE 7.3: Distributions de l'illumination par mesure des médianes locales, $Q1_b$, pour les bases [BQC] et [BQN]

7.1.2 Mesure de la qualité de l'illumination par SVM

En observant les histogrammes de la luminance (voir Chapitre 5, paragraphe 5.4.3) des images de type contrôlé et ceux des images acquises en conditions diverses, nous avons constaté une nette différence de comportement. Pour les images de type contrôlé, ces histogrammes sont plus proches d'une distribution normale alors que ceux des images de type non contrôlé peuvent avoir des formes diverses. La figure 7.4 montre des exemples de ces histogrammes pour les deux types d'images.

Afin de classifier les histogrammes des images de type contrôlé par rapport aux histogrammes des images acquises dans des conditions diverses, nous proposons d'utiliser l'outil de classification par SVM ("Séparateur à Vaste Marge", appelé aussi "Machine à Vecteurs de Support" [75]). En quelques mots, un SVM cherche le meilleur hyperplan séparateur qui minimise le taux d'erreur total (TER) de classification dans un espace de grande dimension. L'apprentissage du SVM a été réalisé sur une base de données dédiée. Pour la partie apprentissage, deux classes d'images de visage sont considérées. La première classe est constituée d'images acquises dans des conditions d'illumination contrôlées et la deuxième classe est constituée d'images acquises dans des conditions d'illumination diverses. L'étiquetage des images a été fait par "NIST" suivant les conditions d'acquisition de

l'image, à savoir qu'une image d'une session contrôlée est considérée comme de bonne qualité et une image acquise dans une session non contrôlée est considérée comme dégradée.

La mesure de qualité considérée est la distance entre l'échantillon à l'hyperplan séparateur calculé par SVM. Cette mesure sera notée Q2 dans la suite du chapitre.

(A) Histogramme de la luminance d'un visage de type contrôlé



(B) Histogramme de la luminance d'un visage de type non contrôlé



FIGURE 7.4: Exemples des histogrammes des images de type contrôlé (A) et de type non contrôlé (B)

Apprentissage du SVM

Des deux bases [BQC] et [BQN], nous choisissons 1000 images de type contrôlé et 1000 images de type non contrôlé pour calculer l'hyperplan séparateur. Le reste des images constituant [BQC] et [BQN], sera utilisé pour optimiser la classification (cette base sera appelée "base de développement").

La luminance de chaque image est extraite, ensuite l'histogramme de cette luminance est calculé. Nous obtenons un vecteur de 255 composantes qui caractérise la distribution de l'illumination sur l'image.

Les noyaux sélectionnés pour cette étude sont :

- noyau linéaire : $k(x, x') = x.x'$

- noyau polynomial : $k(x, x') = (x \cdot x')^d$
- noyau radial (RBF) : $k(x, x') = \exp(-\gamma \|x - x'\|^2)$

Pour les noyaux polynomial et radial, une optimisation des paramètres d et γ est faite sur cette base de développement. La figure 7.5 représente la variation du taux de bonne classification en fonction du rang d du polynôme. La figure 7.6 représente la variation du taux de bonne classification en fonction du coefficient γ du noyau radial.

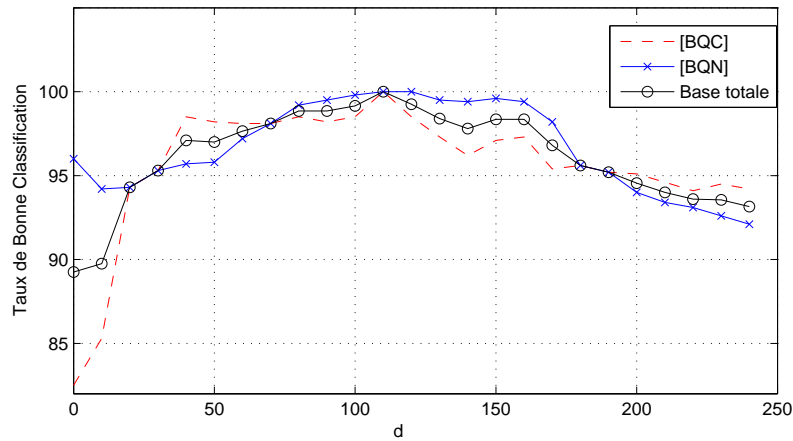


FIGURE 7.5: Variations du taux de classification correcte pour le classifieur SVM à noyau polynomial suivant le rang du polynôme sur la base de développement

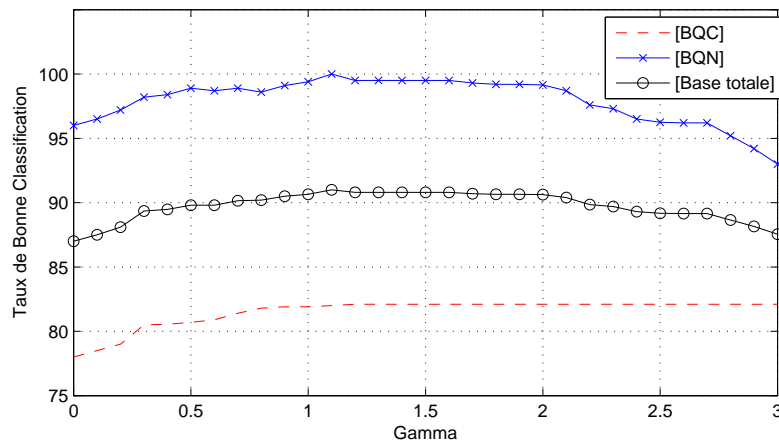


FIGURE 7.6: Variations du taux de classification correcte pour le classifieur SVM à noyau radial sur la base de développement

Le tableau 7.1 présente les meilleurs résultats de classification suivant le type de noyau choisi par rapport à l'hyperplan.

On peut constater que les taux de classification des images non contrôlées atteignent les 100% de bonne classification. Ce taux s'explique par le fait que le SVM cherche à séparer des histogrammes avec des allures quelconques (comme dans le cas des images non contrôlées) par rapport à des histogrammes avec une forme spécifique (le cas des images contrôlées).

Noyau	Linéaire	RBF(1.0)	Polynomial(120)
TBC([BQC])	82,1%	82%	95,5%
TBC([BQN])	100%	100%	100%
TBC(Base totale)	91%	90.5%	97,2%

TABLE 7.1: Taux de Bonne Classification (TBC) pour chaque type de noyau pour les deux bases [BQC] et [BQN]

Pour cette classification, l'étiquetage des deux classes est fait suivant les conditions d'acquisition de l'image. Dans certains cas, une image annotée non contrôlée peut donc avoir une forme d'histogramme qui correspond à un histogramme d'image contrôlée et vice-versa, dans certains cas, les images acquises dans des conditions contrôlées peuvent présenter un déséquilibre d'illumination, produisant ainsi une erreur de classification qui n'a pas lieu d'être.

7.1.3 Mesure de la netteté

Afin de détecter le flou dans les images, plusieurs méthodes sont disponibles dans la littérature. Nous avons testé les deux méthodes les plus classiques : la mesure de netteté par gradient et l'analyse dans le domaine fréquentiel de Fourier.

- la mesure de netteté par méthode du gradient est la différence en moyenne entre le gradient d'une image et la moyenne du gradient de la même image lissée. Si l'image est floue, les deux moyennes seront comparables. En revanche, si l'image est nette, le lissage fait perdre beaucoup de contours et la différence est grande.

La mesure de netteté, qu'on notera $Q3_a$, peut être calculée de la manière suivante :

$$Q3_a = \frac{\text{moyenne}(\text{grad}(\mathbf{I})) - \text{moyenne}(\text{grad}(\mathbf{I} * \mathbf{G}))}{\text{moyenne}(\text{grad}(\mathbf{I}))} \quad (7.1)$$

où I est l'image analysée, $grad$ est l'image du gradient obtenue, par exemple, en appliquant le filtre de Sobel et G est un filtre Gaussien.

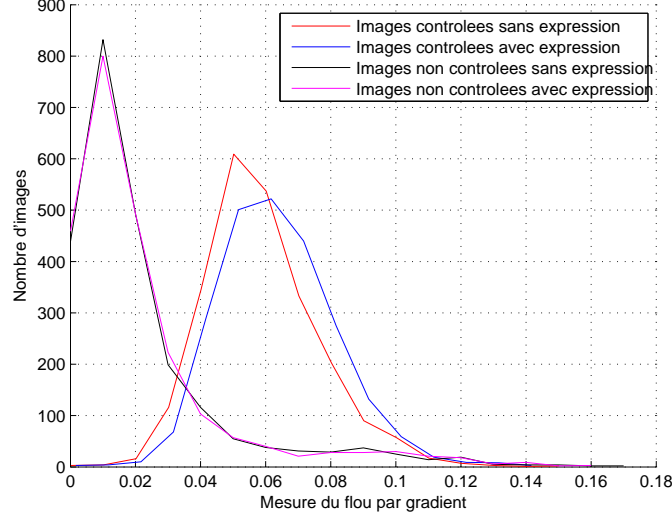


FIGURE 7.7: Distributions des mesures de netteté $Q3_a$ de [BQC] et de [BQN] par méthode du gradient

La figure 7.7 montre clairement une distinction entre les 2 classes (images acquises dans des conditions de mise au point contrôlée (arrière-plan uniforme et distance du visage assez proche de l'objectif de la caméra) et des images acquises dans des conditions diverses (arrière-plan texturé et distance grande entre le visage et la caméra). Ces deux distributions montrent que la mesure du flou est un critère à ne pas négliger pour une classification automatique d'une image de visage.

- le domaine de Fourier est intensivement utilisé comme moyen d'estimer le flou d'une image. En particulier, l'amplitude de la transformée de Fourier est un bon indicateur du niveau de flou dans l'image. En effet, une grande partie de l'information d'une image floue est localisée dans les basses fréquences et inversement, pour une image nette. Ceci se traduit dans le domaine de Fourier par un rapport élevé pour les images floues entre les amplitudes des basses fréquences et la somme totale des amplitudes et inversement, par un rapport faible pour les images nettes. Dans notre étude, nous allons étudier la mesure de qualité $Q3_b$ qui est calculée comme suit :

$$Q3_b = \frac{\sum_{\rho=0}^R F(\rho, \theta)}{\sum_{\rho, \theta} F(\rho, \theta)} \quad (7.2)$$

où $F(\rho, \theta)$ est la transformée de Fourier de l'image en coordonnées polaires et R est le rayon du disque des basses fréquences considérées. Plus la mesure du critère $Q3_b$ est faible, plus l'image est nette car l'information est alors bien étalée entre les hautes et les basses fréquences. A l'inverse, si le critère est élevé (proche de 1), l'image est floue car l'information est en grande partie localisée dans les basses fréquences.

Nous avons considéré une variante de cette mesure qui consiste à mesurer la moyenne de ce critère de netteté sur des imagerie du visage. La mesure de netteté $Q3_c$ s'écrit sous la forme :

$$Q3_c = \sum_{i=1}^N w_i * \frac{\sum_{\rho=0}^R F_{Zi}(\rho, \theta)}{\sum_{\rho, \theta} F_{Zi}(\rho, \theta)} \quad (7.3)$$

Le poids w_i , de la zone Zi , a été optimisé sur la base de développement afin de maximiser la séparation entre les deux classes pré-étiquetées. La figure 7.8 montre le choix du découpage d'une image de visage et la méthodologie de la mesure du critère.

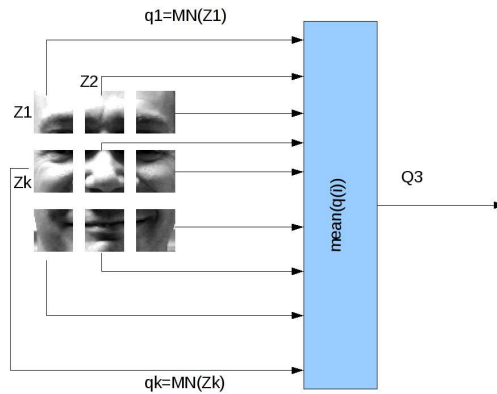


FIGURE 7.8: Mesure du flou par approche fréquentielle locale pondérée

Les figures 7.9 et 7.10 montrent une distinction dans les deux distributions des 2 bases selon que l'on utilise $Q3_b$ ou $Q3_c$.

L'approche fréquentielle locale pondérée montre un net avantage sur l'approche fréquentielle globale classique. Ceci s'explique par la nature de l'image du visage. Le visage présente, en effet, une grande variation en terme de

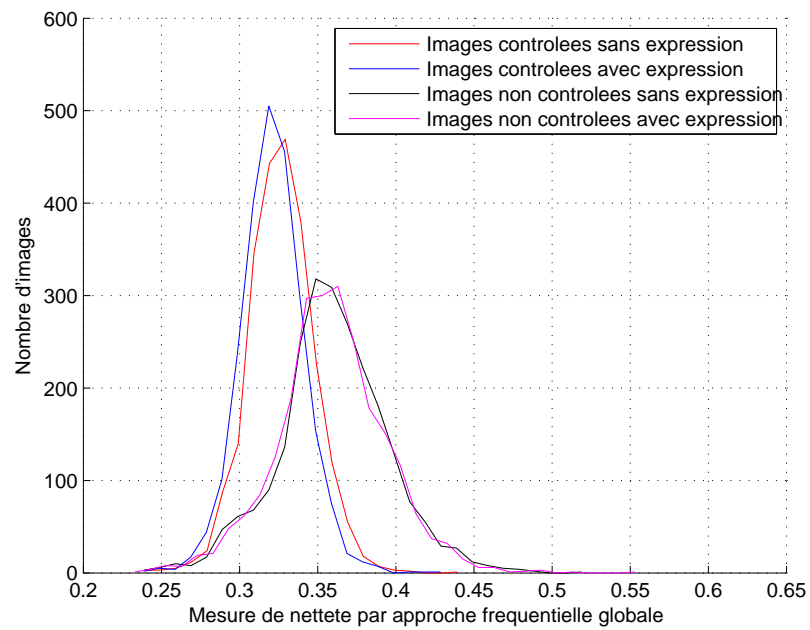


FIGURE 7.9: Distributions des mesures de netteté $Q3_b$ de [BQC] et de [BQN] par l'approche fréquentielle globale

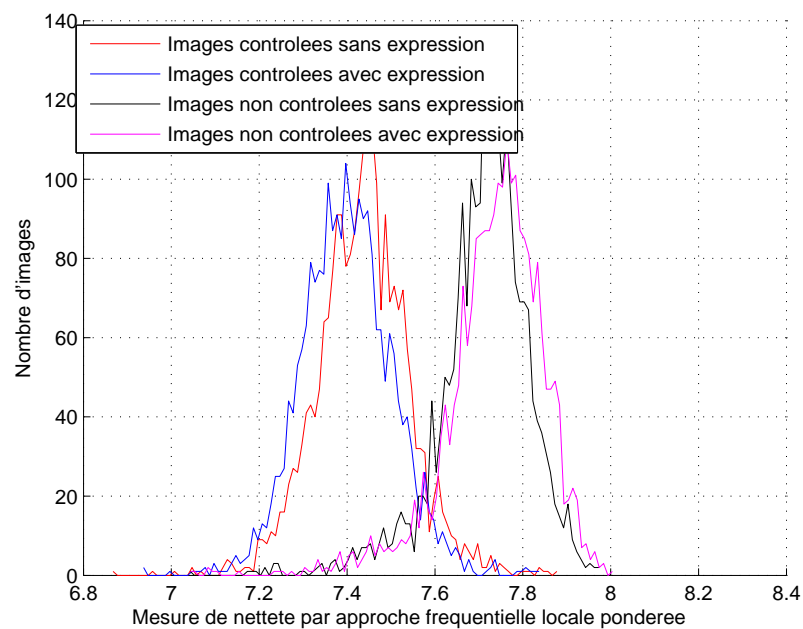


FIGURE 7.10: Distributions des mesures de netteté $Q3_c$ de [BQC] et de [BQN] par l'approche fréquentielle locale pondérée

texture suivant différentes zones. Par exemple, certaines zones du visage comme le front ou les joues ont une texture lisse qui peut altérer la mesure de netteté par l'approche fréquentielle globale.

Mesure de netteté	AG%	AFG	AFLP
TEC	10,1%	22,5%	7%

TABLE 7.2: Taux d'Erreur de Classification (TEC) suivant la méthode de mesure (AG : Approche Gradient, AFG : Approche Fréquentielle Globale, AFLP : Approche Fréquentielle Locale Pondérée)

Pour le reste de l'étude, nous considérons la mesure de qualité basée sur l'approche fréquentielle locale pondérée car elle présente les meilleures performances de classification.

7.1.4 Classification par fusion des mesures de qualité

Nous allons nous positionner, dans ce travail, dans une optique de classification de la qualité des images de visage par les mesures déjà étudiées. Plusieurs méthodes s'offrent à nous. Nous pouvons citer deux types d'approches : les méthodes de classification par apprentissage supervisé et les méthodes par apprentissage non supervisé [9, 75]. Dans notre étude, nous nous sommes intéressés exclusivement à l'apprentissage non supervisé qui va nous permettre de catégoriser les bases de façon automatique sans nécessiter de pré-étiquetage *a priori*. Le seul paramètre à fixer reste le nombre de classes à chercher.

Dans la panoplie des algorithmes de classification par apprentissage non supervisé, nous pouvons citer la méthode des nuées dynamiques, le regroupement hiérarchique, l'algorithme EM, l'analyse en composantes principales et également la méthode des k-moyennes. Nous n'allons pas nous attarder ici sur le principe de fonctionnement de ces méthodes [75].

Nous présentons ici nos résultats de classification avec la méthode des k-moyennes¹.

Pour la recherche des paramètres des centroïdes, nous avons utilisé une partie des deux bases [BQC] et [BQN] : 1000 images de [BQC] et 1000 images de [BQN].

1. L'algorithme des k-moyennes est également dénommé "algorithme des centres mobiles" ou "centroïdes". L'objectif est de segmenter les données en k groupes, k étant fixé *a priori*. L'idée de cet algorithme est très intuitive et, de fait, cet algorithme a été réinventé à plusieurs reprises. Il en existe de nombreuses variantes, en particulier l'algorithme bien connu des "nuées dynamiques".

Pour chaque échantillon x de la base, nous notons le vecteur $Q_x = \{Q1, Q2, Q3\}$, avec $Q1, Q2$ et $Q3$: les trois mesures de qualité, étudiées dans la section précédente.

Sur cette base d'apprentissage, nous avons effectué plusieurs tests afin de déterminer le nombre optimal des classes de qualité (centroïdes). La méthode se base sur la variation de l'inertie intra-classes en fonction du nombre de classe.

Définition : *L'inertie intra-classes d'un ensemble est définie comme suit :*

$$I_w = \sum_{i=1}^K \left(w_i * \sum_{j=1}^{N_i} d(x_j, G_i) \right) \quad (7.4)$$

avec :

- K le nombre de classes
- w_i le poids du groupe G_i , $w_i = N_i/N$ où N est le total des échantillons de l'ensemble et N_i le cardinal du groupe G_i
- $d(x_j, g_i)$ la distance euclidienne de l'échantillon x_j au centre de gravité du groupe G_i .

Afin de déterminer le nombre minimal de groupes susceptibles d'exister dans un ensemble, on peut tester plusieurs valeurs de K en exécutant plusieurs fois l'algorithme avec des K croissants. Pour chaque valeur de K , on calcule l'inertie intra-classes. Cette valeur décroît quand K augmente. En faisant un graphique représentant l'inertie intra-classes en fonction de K , on voit la bonne valeur de K : c'est celle à partir de laquelle I_w ne décroît plus de façon significative.

La figure 7.11 montre l'évolution de l'inertie intra-classes en fonction du nombre de classes choisi.

Nous constatons une stabilisation de l'inertie à partir de 3 centroïdes, sachant que la différence entre 2 ou 3 centroïdes n'est pas énorme. Par la suite, nous fixons le nombre de classe à 2. Afin de classer un échantillon de test dans un groupe (en terme de qualité), nous affecterons cet échantillon à la classe la plus proche (en terme de distance par rapport au centre de gravité des différentes classes de qualité). Nous définissons pour la suite la mesure de qualité globale q d'un échantillon par :

$$q = \frac{d(Q_x, Q_{g_2}) - d(Q_x, Q_{g_1})}{d(Q_{g_1}, Q_{g_2})} \quad (7.5)$$

avec $d(Q_x, Q_{g_j})$ la mesure de distance entre le vecteur de qualité de l'échantillon x et le vecteur de qualité du centre de gravité de la classe j . Cette mesure présente

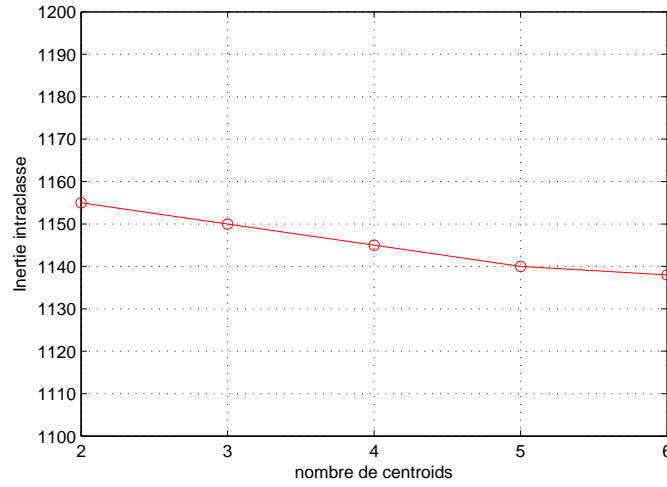


FIGURE 7.11: Évolution de l'inertie intra-classes en fonction du nombre des centroïdes pour la base d'apprentissage (1000 images de [BQC] + 1000 images de [BQN])

deux états de l'échantillon :

$$q > 0 \Rightarrow x \in g1$$

$$q < 0 \Rightarrow x \in g2$$

Par la suite, on pose $g1$ la classe des échantillons de bonne qualité et $g2$ la classe des images dégradées. Plus q est grand, plus l'échantillon est considéré de bonne qualité et inversement, plus q est petit, plus l'échantillon est considéré de qualité dégradée.

La figure 7.12 nous montre les distributions de la mesure q pour les images de la base d'apprentissage [[BQC] + [BQN]].

Nous pouvons noter que les 2 bases [BQC] et [BQN] sont bien séparées mais qu'il existe tout de même un recouvrement.

Remarques :

- la mesure q n'est valable que pour les problèmes à deux classes.
- nous avons effectué une autre étude de classification par fusion des mesures de qualité Q1, Q2 et Q3 en se basant sur les SVM, mais cette étude a montré des limites de généralisation sur des bases de validation en particulier sur la base BANCA, dues à un manque d'échantillon d'apprentissage.

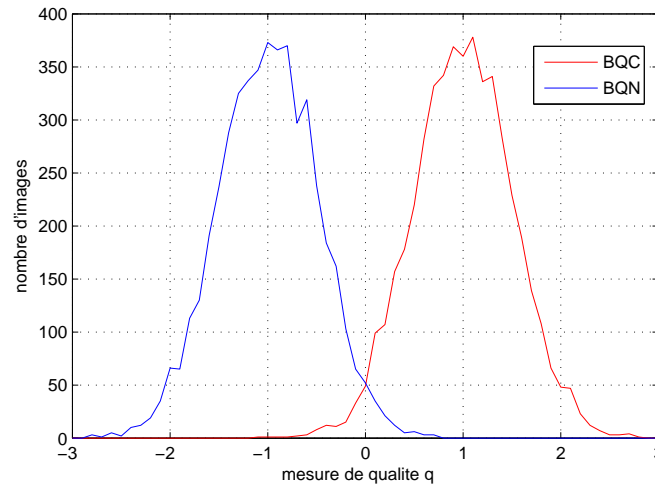


FIGURE 7.12: Distribution de la mesure q pour les images de la base de développement $[[BQC] + [BQN]]$

7.1.5 Validation des mesures de qualité sur la base de développement

Validation des mesures sur la base BANCA

La base BANCA comme présentée dans le chapitre 3 présente 3 scénarios d'acquisition : une acquisition en mode contrôlé en illumination et en mise au point, une acquisition en mode dégradé (*webcam*) et en illumination non contrôlée et un dernier scénario d'acquisition en mode non contrôlé en illumination. Les exemples de la figure 7.13 montrent la différence entre les différents scénarios.

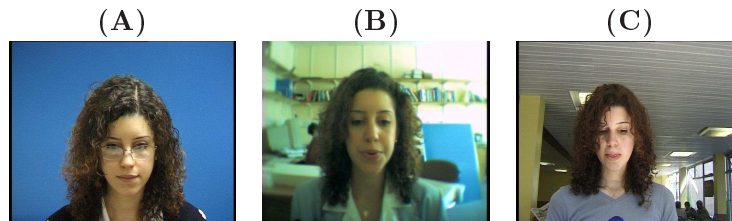


FIGURE 7.13: Exemples d'images de type contrôlé (A), de type dégradé (B) et de type non contrôlé (C) de la base de données BANCA

Nous divisons la base BANCA en deux sous-bases d'image de visage. La première sous-base est composée des sessions [s01–s04] et sera considérée comme l'ensemble de la classe des images de bonne qualité (Classe 1). La deuxième sous-

base sera composée des images des sessions [s05–s12] et sera considérée comme l'ensemble de la classe des images de qualité dégradée (Classe 2).

Un ré-apprentissage des centroïdes est nécessaire dans le cas de base de BANCA. Ce ré-apprentissage est nécessaire dû à la différence de qualité des capteurs. Pour cela, nous avons considéré la base de développement de la base BANCA, qui contient des échantillons de 30 personnes ne participant pas à l'évaluation. Ces échantillons proviennent des sessions s01, s05 et s09. Pour la base FRGCv2, la même caméra a été utilisée pour les différentes sessions d'acquisition, alors que pour BANCA, le même type de caméra a été utilisé pour les sessions [s01-s04 et s09-s12] et une webcam a été utilisée pour l'acquisition des sessions [s05-s08].

Le tableau 7.3 donne les résultats de bonne classification des différentes classes avec notre mesure de qualité Q définie précédemment.

Sessions	Classe 1 [s01–s04]	Classe 2 [s05–s12]
TCC	85%	99.2%

TABLE 7.3: Taux de Classification Correcte (TCC) pour les deux classes sélectionnées de la base BANCA avec notre mesure q

Les résultats de classification de la base BANCA confirment les résultats obtenus dans la première partie, à savoir qu'il est plus facile de détecter des images de type dégradé que des images de type contrôlé avec nos critères de qualité.

7.2 Qualité et stratégie de reconnaissance

Dans la majorité des protocoles d'évaluation des bases de visage (voir chapitre 3), on constate une mise en place des protocoles de test suivant les conditions d'acquisition des échantillons, afin d'étudier la robustesse des algorithmes suivant certain type de dégradation.

Ainsi, pour la base FRGCv2, deux protocoles d'évaluation de visage 2D sont mis en place : l'expérience 1 et l'expérience 4. Ces deux protocoles se basent sur la qualité des échantillons en test et en référence. Aussi pour la base BANCA, les différents protocoles d'évaluation proposés sont basés sur la qualité d'acquisition des échantillons. Récemment, les protocoles de tests mis en place pour la campagne d'évaluations de visage 2D pour les campagnes FRVT2006 et MBGC se basent sur la même stratégie.

Dans cette partie du chapitre, nous présentons des résultats qui confirment

cette différence de performance. Nous proposons également des stratégies de reconnaissance basées sur les mesures de qualité des images du visage et nous validons cette approche sur la base BANCA.

7.2.1 Protocoles et bases de données

La base de données utilisée dans cette étude est la base FRGCv2. Elle sera découpée dans le but d'étudier plusieurs types de scénarios. Le premier type de scénarios est l'étude du comportement du classifieur lorsque les images de référence sont de type passeport (suivant les mesures des qualités déjà étudiées). Le deuxième type de scénarios est dédié à l'étude du comportement du classifieur lorsque les images de référence sont dégradées.

Pour le premier type de scénario, plusieurs protocoles sont mis en place suivant la nature des images de test :

- Test1 : images de test de type passeport
- Test2 : images de test acquises dans un environnement non contrôlé et sans expression
- Test3 : images de test de bonne qualité (illumination, netteté) et avec expressions
- Test4 : images de test acquises dans un environnement non contrôlé et avec expressions

Les images de référence sont de type passeport².

Découpage de la base de données

La base FRGCv2 se compose de 2 parties distinctes (développement et validation, voir Fig 7.14). La partie développement est dédiée à la construction des espaces de réduction³ et à l'optimisation des systèmes. Elle contient les enregistrements de 222 personnes et 12736 images au total, dont ~4200 sont acquises en environnement non contrôlé et ~8500 en conditions d'illumination et de mise au point contrôlées.

La partie validation de la base FRGC contient les enregistrements de 466 personnes (dont 222 sont dans la base de développement). Cette partie de la base est elle-même divisée en 2 parties (suivant la nature des images) : une première

2. Voir l'introduction du chapitre pour une définition

3. Les espaces de réduction sont les espaces relatifs à l'algorithme de reconnaissance par DLDA utilisé dans cette étude.

partie contient 16028 images de type contrôlé (en illumination et en netteté), dont la moitié sont avec une expression neutre, et 8014 images acquises en conditions non contrôlées dont, aussi, la moitié sont avec une expression neutre.

Pour cette étude, les 220 personnes de la base de développement seront utilisées pour créer 2 types d'espaces de réduction.

- P0 : 10 images par personne, de type contrôlé.
- P1 : 20 images par personne, dont 10 sont de type contrôlé et 10 de type non contrôlé.

Pour la sélection de la base de référence (nommé [B0] issue de la base de validation), la mesure de qualité définie dans le paragraphe 7.1.4 sera exploitée. Parmi les 8014 images contrôlées et sans expression, 6414 images appartenant à 441 personnes sont sélectionnées, suivant des mesures de qualité très strictes. Les 4 meilleures images (du point de vue mesure de qualité) de chaque personne sont ensuite utilisées comme images de référence. La base B0 est alors constituée de 884 images de référence de 441 personnes.

Afin d'effectuer le protocole Test1, 4 autres images, type contrôlé, des 441 personnes sont sélectionnées aléatoirement à partir des 5530 images de type passeport restantes. Cette base est nommée [BT1] et elle est constituée aussi de 884 images.

Pour le protocole Test2, la base de test [BT2] est constituée des 441 personnes avec 4 images par personne, les images sont de type non contrôlé et sans expression. Elles sont choisies aléatoirement à partir des 4007 images non contrôlées et sans expression.

Pour le protocole Test3, la base de test [BT3] est constituée des 441 personnes avec 4 images par personne. Les images sont sélectionnées à partir des 4007 images contrôlées et avec expressions.

Pour le protocole Test4, la base de test [BT4] est constituée des 441 personnes avec 4 images par personne. Les images sont sélectionnées à partir des 4007 images non contrôlées et avec expressions.

Le graphe de la figure 7.14 résume les différents découpages et la constitution des bases de test à partir de la base originale de FRGCv2.

Remarque : Il faut savoir que les 222 personnes de la base de développement sont incluses dans la base de validation mais que les données de la base de validation et ceux de la base de développement sont complètement différentes.

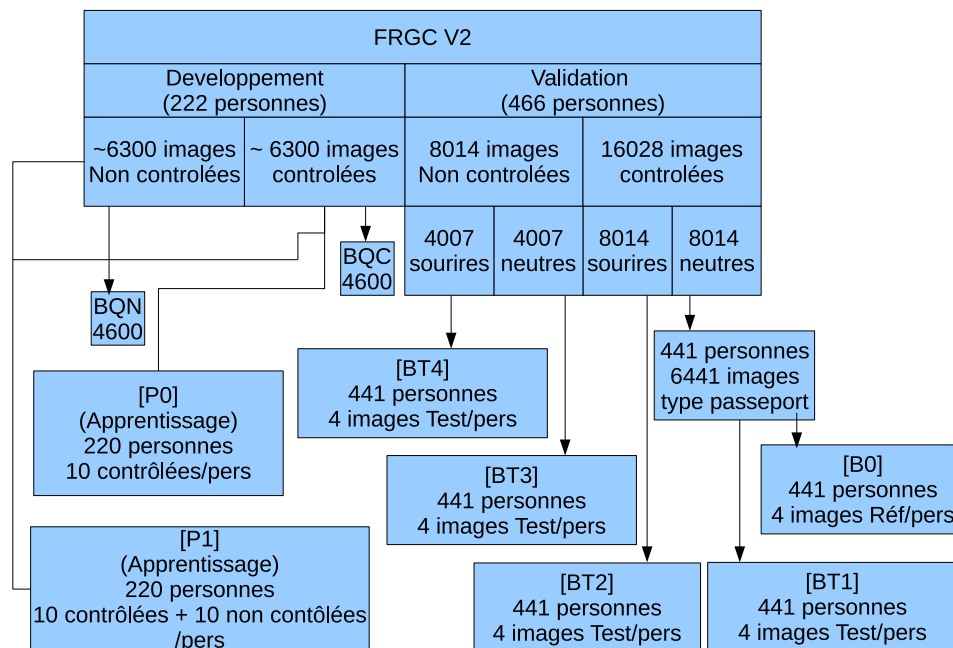


FIGURE 7.14: Découpage de la base FRGCv2

7.2.2 Scénario : références Passeport

L'algorithme utilisé dans cette étude est la DLDA (voir chapitre 4 et chapitre 6), appliqué aux transformations de Gabor de l'image de visage. Pour cette étude, 2 types de pré-traitement sont considérés : l'égalisation d'histogramme et le lissage anisotropique.

Le premier type de pré-traitement est appliqué aux images, considérées comme acquises dans des conditions contrôlées. Le lissage anisotropique est appliqué lorsque l'image de test est de type non contrôlé.

Test1

Les images de [B0] et de [BT1] sont des images de type contrôlé (illumination et flou), une simple égalisation d'histogramme sera appliquée aux images en entrée, avant le filtrage de Gabor.

Dans une première expérience, la réduction de dimension par DLDA sera appliquée en utilisant l'espace P0 (l'espace de réduction est appris sur des images contrôlées seulement). Dans une deuxième expérience, l'espace P1 (appris sur des

images contrôlées + des images non contrôlées) sera utilisé.

La figure 7.15 montre clairement la différence de performances entre les 2 expériences. Le fait d'utiliser un espace de réduction, appris sur des données propres et des données bruitées, dégrade de façon significative les performances. Le taux de vérification à 0.1% de FAR passe de 96,2% de reconnaissance en utilisant l'espace P0 à 93,1% avec l'espace P1 avec des intervalles de confiance de $\pm 0,5\%$ (voir chapitre 3) pour plus de détails sur la mesure de confiance. Le résultat de

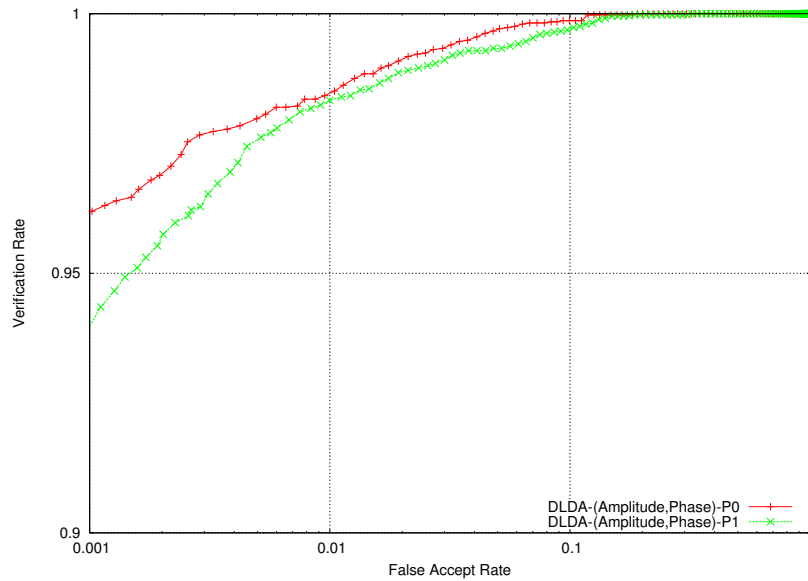


FIGURE 7.15: Courbes ROC de Test1

96,2% de VR@0.1% de FAR de ce protocole donne une idée sur le maximum de performance qu'on peut atteindre avec l'algorithme DLDA dans les meilleures conditions : un espace de réduction appris sur des données propres, des données de tests de qualité optimale et des visages avec des expressions neutres.

Test2

Ce protocole a pour but l'étude des performances de l'algorithme de reconnaissance DLDA, dans des conditions dégradées d'acquisition de l'image de test. Les deux types d'expériences, du protocole Test1, sont repris afin d'étudier aussi l'influence des espaces de réduction.

L'influence de l'espace de réduction est d'autant plus importante dans ce protocole, étant donné la nature des images de test. Lorsque l'espace de réduction est appris sur des images de type contrôlé, il ne prend pas en compte les dif-

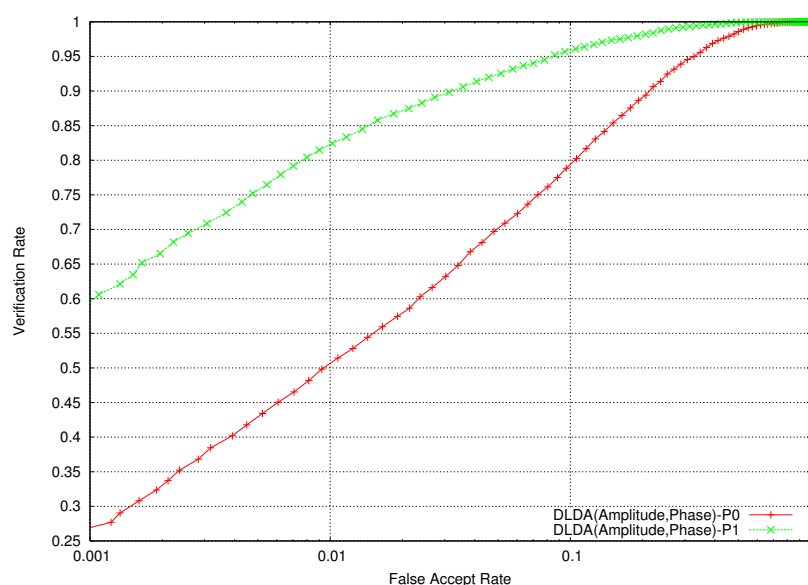


FIGURE 7.16: Courbes ROC de Test2

férentes variations qui peuvent subvenir sur une image de type dégradé, et ceci même en utilisant un pré-traitement assez complexe pour réduire l'influence de l'illumination. Dans ce type de condition de test, l'apprentissage des dégradations est indispensable. Le taux de vérification à 0.1% de FAR passe de 59,6% [$\pm 1.5\%$] avec P1 à 27,4% [$\pm 2.5\%$] avec P0.

Test3

Le Test3 est consacré à l'étude de l'influence des variations d'expression dans la phase de test sur les performances du même algorithme. Il faut noter que dans la base d'apprentissage, une partie des visages comporte des expressions, en l'occurrence des légers sourires. L'espace de réduction considéré est P0.

La base BT3 utilisée est une base d'images contrôlées en illumination et en netteté. Le taux de vérification à 0.1% de FAR passe de 96,2% de reconnaissance (Test1) à 95,8% pour le Test3 avec des intervalles de confiance de $\pm 0,5\%$. Les courbes ROC de la figure 7.17 montrent que l'influence d'une légère expression n'est pas vraiment très significative et que les performances de l'algorithme restent stables.

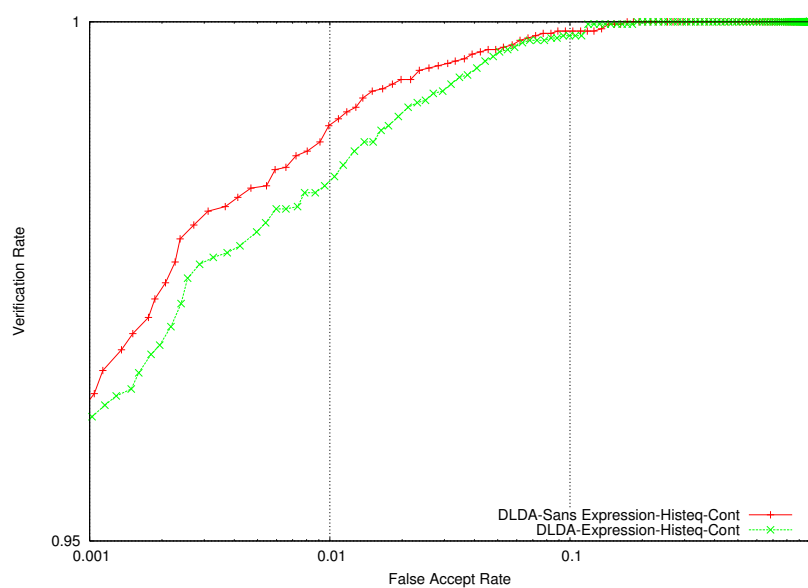


FIGURE 7.17: Courbes ROC de Test3

Test4

Le Test4 est équivalent au Test3, appliqué aux conditions dégradées d'acquisition. L'espace de réduction est construit à partir de P1, et le pré-traitement reste le lissage anisotropique.

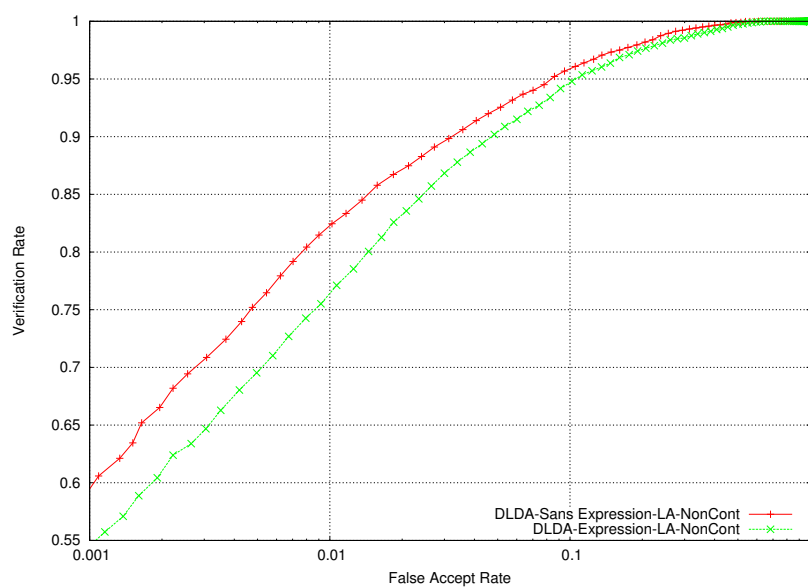


FIGURE 7.18: Courbes ROC de Test4

Le taux de vérification à 0.1% de FAR passe de 59,6% [$\pm 1.4\%$] à 54,7%

$[\pm 1.6\%]$. Les courbes ROC de la figure 7.18 montrent que l'influence d'une légère expression est un peu plus sensible dans les conditions dégradées par rapport aux conditions non dégradées. Cependant, la dégradation des résultats reste assez faible, comparée à celle induite par des variations de l'illumination et de la netteté de l'image.

Test5 ou choix de la stratégie

Dans le Test5, le but est de démontrer qu'il est plus judicieux d'adapter le classifieur à qualité du test, plutôt que d'utiliser le même classifieur pour n'importe quel type de test (bonne qualité contre bonne qualité ou bonne qualité contre image dégradée). Cette idée semble simple et intuitive, ce protocole permet de valider cette stratégie. Pour cela, nous considérons [B0] comme une base de référence et les bases [BT1] et [BT2] sont mélangées pour créer la base de test.

Une première expérience consiste à appliquer l'algorithme de la DLDA en se basant sur l'espace construit avec P0 et en utilisant une égalisation d'histogramme comme pré-traitement.

Une deuxième expérience consiste à utiliser l'espace construit à partir de P1 en utilisant le lissage anisotropique comme pré-traitement.

La troisième expérience consiste à utiliser le meilleur espace pour le type de test considéré. En d'autres termes, si le test est une comparaison entre une image de type contrôlé contre une image de type contrôlé, l'espace P0 sera utilisé comme espace de réduction. Dans le cas d'une comparaison entre une image de type contrôlé et une image de type non-contrôlé, l'espace de réduction construit à partir de P1 sera appliqué.

La différence de performance [figure 7.19], suivant la stratégie choisie, démontre l'importance du choix de l'espace de réduction suivant la nature du test. Le taux de vérification à 0.1% de FAR passe de 58,2% $[\pm 1.4\%]$ pour la première expérience à 78.4% $[\pm 1.2\%]$ pour la deuxième et finalement à 84,9% pour une stratégie basée sur l'espace adapté aux types de tests.

Test6

Pour le Test5, le choix de l'espace de réduction pour la troisième expérience était basé sur une connaissance *a priori* du type de test (étiquetage fourni par NIST). Dans ce protocole (Test6), j'exploite les mesures de qualité que j'ai défini

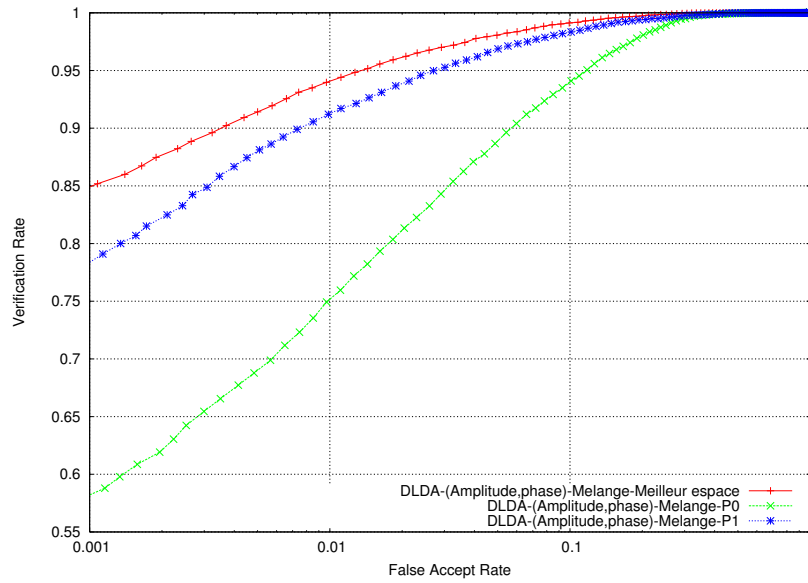


FIGURE 7.19: Courbes ROC de Test5

sur l'image test pour décider de la nature du test de comparaison ⁴.

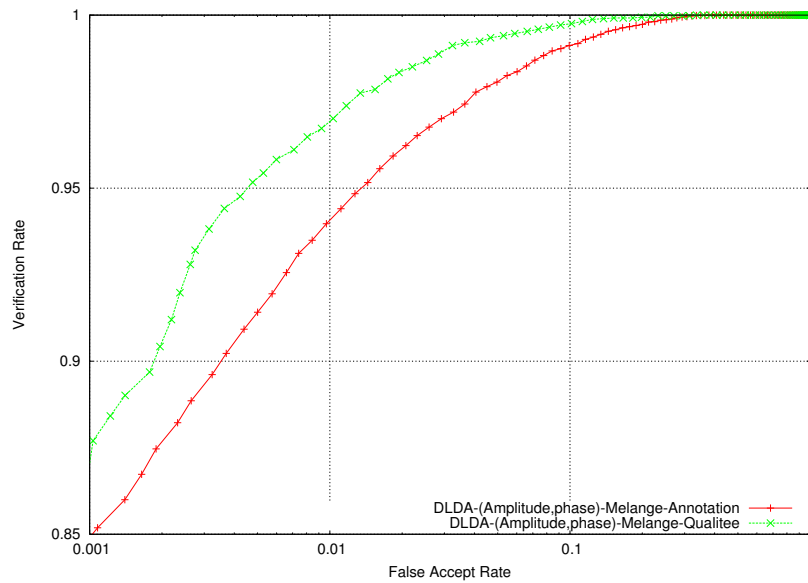


FIGURE 7.20: Courbes ROC de Test6

La figure 7.20 montre une légère amélioration des résultats grâce à l'usage des mesures de qualité. Le taux de vérification à 0.1% de FAR passe de 84,9% [$\pm 1.1\%$]

4. Comme expliqué dans la section précédente : même si les annotations des images de visage restent dépendantes du type de l'acquisition, cette annotation peut ne pas correspondre à la réalité (voir figure 7.12)

avec une annotation manuelle des images à 86.9% [$\pm 0.9\%$].

Même si cette amélioration n'est pas très significative, elle démontre que les mesures de qualité choisies ont une influence sur les performances et qu'elles contribuent au choix de la stratégie pour l'application du classifieur. Ce qui est intéressant est surtout le fait de pouvoir exploiter ces mesures de qualité pour le cas où on ne disposerait pas d'étiquetage *a priori* des données comme dans des conditions réelles d'application. La stratégie dont l'efficacité est démontrée sur les images annotées à la main est corroborée par l'utilisation de la mesure automatique de qualité.

7.2.3 Scénario : références dégradées

Ce scénario est choisi pour sa ressemblance à l'application de reconnaissance du visage à partir de deux séquences vidéo de surveillance. Les images de référence sont des images de type dégradé et les images de test sont de qualité diverse.

Les images de la base référence [BT3] étant de qualité dégradée, nous considérons l'espace de réduction construit à partir de P1 comme espace de projection et le pré+-traitement par lissage anisotropique sera considéré.

Test7

Pour ce premier protocole du nouveau scénario, on considère la base [BT4] comme base de test. Le modèle de chaque personne est constitué d'une seule image de référence.

La figure 7.21 montre les résultats du Test7. Un taux de vérification de 75% à 0.1% de FAR peut sembler non conforme aux attentes car les performances sont très correctes vu la nature dégradée des images des références et celle des tests. En analysant les distributions des histogrammes du Test2 et celui du Test7 [Figure 7.22], nous pouvons constater un comportement stable des scores inter-classes alors que les scores intra-classes sont plus ressemblants dans le cas d'images de visage de nature semblable.

Test8

Dans les protocoles déjà présentés, le modèle d'une personne est constitué par une seule image de référence. Dans le protocole du Test8, le modèle de référence est constitué des 4 images de référence d'une personne. La distance minimale

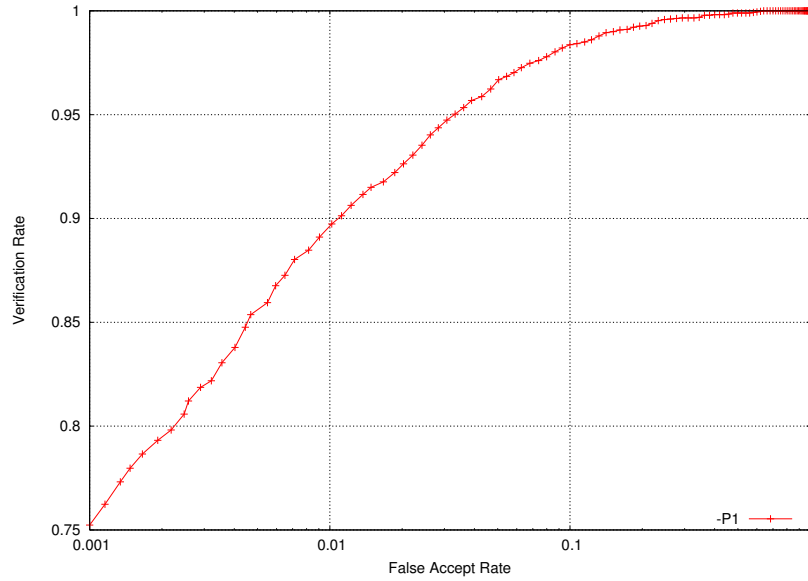


FIGURE 7.21: Courbes ROC de Test7

entre l'image de test et les 4 images de référence est considérée comme le score du test. Le taux de VR passe de 75% pour le Test7 à 78.1% en utilisant un modèle à plusieurs références. Les résultats sont donnés par les courbes ROC de la figure 7.23.

Protocole : Test9

Nous reprenons le même protocole que celui du Test8 mais la distance entre l'image de test et les 4 images de référence est calculée comme une somme pondérée par les mesures de qualité de chaque image de référence. Dans ce protocole, la mesure de qualité de l'image de test n'est pas considérée. Le taux de VR à 0.1% de FAR est de 80.2% en utilisant un modèle à plusieurs références. Les résultats sont donnés par les courbes ROC de la figure 7.23.

Protocole : Test10

Dans ce protocole, nous considérons les modèles référence et test d'une personne formés chacun des 4 images des visages disponibles.

Remarque : Afin d'intégrer la mesure de qualité (Équation (7.5)), comme poids de pondération pour les distances étudiées par la suite, une modification de la mesure de la qualité q est nécessaire. Et afin de donner un poids proche de 1 pour

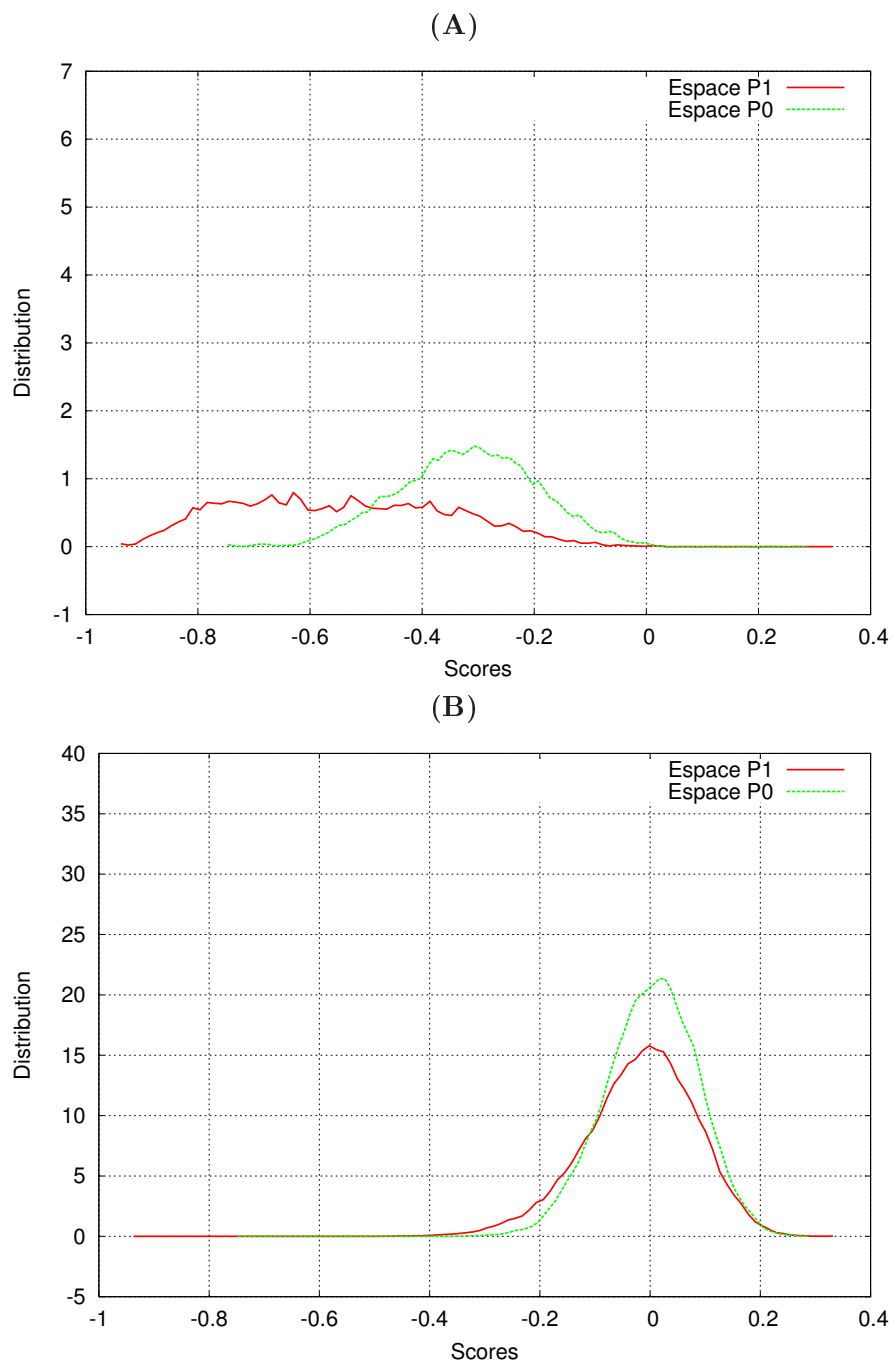


FIGURE 7.22: Distribution des scores intra-classes (A) et Distribution des scores inter-classes (B)

les images de bonne qualité et un poids proche de 0 pour les images de qualité dégradée, nous considérons cette fonction :

$$\begin{aligned}
f(x) &= 1 - \frac{d(Q_x, Q_{g_1}) - d(Q_x, Q_{g_2})}{d(Q_{g_1}, Q_{g_2})} \\
q &= f(x) \text{ if } 0 \leq f(x) \leq 1 \\
q &= 0 \text{ if } f(x) < 0 \\
q &= 1 \text{ if } f(x) > 1
\end{aligned} \tag{7.6}$$

Dans ce Test10, nous avons comparé 3 distances :

- D1 : le minimum des 16 distances (4 images référence contre 4 images test),
D1 = $\min(d_i)$ avec d_i le score de comparaison $i = \{1 - 16\}$.
- D2 : le minimum des 4 distances pondérées (4 images référence contre 1 image test)

$$D2 = \min_t \left(\sum_{i=1}^4 q_r^i * d_i \right) \tag{7.7}$$

avec q_r^i la mesure de la qualité de l'image de référence et d_i le score de comparaison i .

- D3 : la distance totale pondérée par la qualité des échantillons (4 images référence contre 4 images de test)

$$D3 = \sum (w_t^i * q_r^i * d_i) \tag{7.8}$$

avec q_t^i la mesure de la qualité de l'image de test et d_i le score de comparaison i .

Les résultats du Test10 sont illustrés dans la figure 7.23 pour les différentes distances choisies. On constate une amélioration des performances entre l'utilisation de plusieurs images par modèle pour la référence et le test et l'utilisation d'une seule image par modèle de test. Ce résultat confirme aussi des résultats bien connus de Phillips sur la base FRGCv2 [69], avec l'amélioration des performances de la PCA entre l'expérience 1 (utilisant une seule image référence par personne) et l'expérience 2 (utilisant 4 images de référence par personne).

Ce qui est intéressant dans ces résultats est l'amélioration significative apportée par la distance D3 par rapport à l'approche classique utilisant la distance D1. Les résultats passent de 82% [$\pm 0.9\%$] de VR@0.1 de FAR à 88.3% [$\pm 0.7\%$]. Ceci confirme encore que l'approche basée sur les mesures de qualité étudiées dans la première partie du chapitre est très robuste.

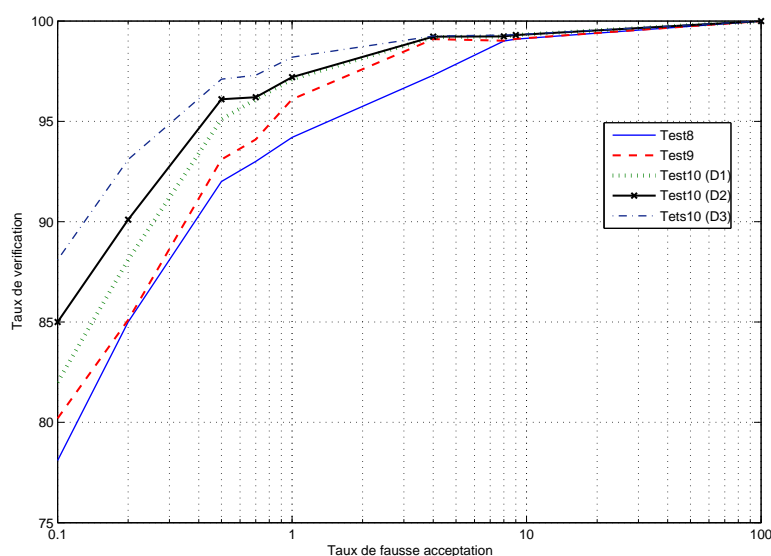


FIGURE 7.23: Courbes ROC des Tests 8-9 et 10 (D1, D2 et D3)

7.2.4 Récapitulatif des tests

En résumé, plusieurs tests ont été effectués afin de valider nos approches de reconnaissance. Pour le premier scénario qui consiste à comparer des images de référence de type contrôlé contre des images de test de qualités diverses, nous avons exploité la classification par mesure de qualité, étudiée dans la première partie du chapitre, afin de choisir la meilleure approche de comparaison. La deuxième stratégie est présentée dans le cadre d'une reconnaissance de visage lorsque les images de références et de test sont de type dégradé. Dans ce cadre, nous avons une connaissance *a priori* de la qualité des images utilisées et nous avons opté alors pour une stratégie de fusion par pondération des mesures de qualité car elle améliore sensiblement les performances.

Le tableau 7.4 récapitule les différences entre les scénarios et les protocoles ainsi que les résultats des divers tests.

Scénarios	Test	N Ref/mod	N Test/mod	pré-trait.	Projection	Résultats
référence passeport	1	1	1	EH	P0	96,2%
	2	1	1	LA	P1	59,6%
	3	1	1	EH	P0	95,8%
	4	1	1	LA	P1	54,7%
	5	1	1	EH/LA	P0/P1	84,9%
	6	1	1	EH/LA	P0/P1	86,9%
référence dégradé	7	1	1	LA	P1	75%
	8	4	1	LA	P1	78.1%
	9	4	1	LA	P1	80.2%
	10	4	4	LA	P1	82%
		4	4	LA	P1	85%
		4	4	LA	P1	88.3%

TABLE 7.4: Tableau récapitulatif des divers scénarios et protocoles avec les résultats des performances donnés en VR@0.1% de FAR (EH : Égalisation d’Histogramme, LA : Lissage Anisotropique). La mesure de distance est le cosinus. P0 : Espace de projection construit avec des images de type contrôlé, P1 : Espace de projection construit avec un mélange d’images de type contrôlé et non contrôlé, Nref/mod : nombre d’images de référence par modèle, NTest/mod : nombre d’images de test par modèle.

7.3 Validation de la stratégie sur la base BANCA

Nous montrons ici la validité de la stratégie basée sur la pondération par les mesures de qualité sur la base BANCA. Les protocoles P et G de la base sont étudiés [4]. L'intérêt de ces deux protocoles est qu'ils correspondent aux protocoles construits pour notre étude à savoir, pour le protocole P, nous avons pour chaque modèle de référence 5 images de qualité contrôlée et le modèle de test peut avoir diverses qualités. Pour le protocole G, les modèles de référence et de test sont aussi composés de 5 images de visage mais les modèles de référence peuvent être de qualités diverses.

Afin de pouvoir comparer avec les résultats de l'état de l'art, nous avons utilisé le point de fonctionnement WER (*Weighted Error Rate*). Les détails du calcul des performances ainsi que les résultats des autres évaluations peuvent être retrouvés dans le chapitre 3 paragraphe 3.4.2 et l'article de Messer [55].

La table 7.5 présente les résultats de notre stratégie (basée sur la mesure de qualité) et en la comparant à la stratégie sans exploitation des mesures de qualités.

Protocole	Distance	WER(0.1)		WER(1)		WER(10)		Av.WER%
		G1	G2	G1	G2	G1	G2	
P	D1	4,3	4,5	10,30	11,75	3,80	4,53	6,53
	D3	3,5	3,10	7,30	6,50	3,25	3,30	4,49
G	D1	2,49	2,66	4,20	5,63	2,85	2,32	3,35
	D3	1,4	1,6	2,05	2,2	1,6	1,52	1,72

TABLE 7.5: Résultats de la stratégie de reconnaissance sur la base BANCA, Protocole P et G, utilisant l'algorithme de reconnaissance DLDA appliqué aux représentations Gabor du visage.

Les résultats du tableau montrent clairement une amélioration des résultats en s'appuyant sur les mesures de qualité des différents échantillons et confirme que la stratégie adoptée est robuste et généralisable sur d'autres bases d'évaluation.

7.4 Conclusion

Afin d'étudier l'apport des stratégies de reconnaissance basées sur la qualité, nous avons utilisé un système de reconnaissance de visage classique basé sur la DLDA, appliqué aux représentations d'amplitude et de phase de Gabor, le système a été présenté en détail dans le chapitre 6. Ce chapitre commence par une étude

de plusieurs mesures de qualité comme la netteté et l'illumination. Trois mesures de qualité ont été adoptées et nous avons validé ces mesures sur les deux bases FRGCv2 et BANCA. Ces deux bases présentent des caractéristiques intéressantes pour notre étude, à savoir des conditions d'acquisition diverses et des qualités différentes.

Une deuxième partie de ce chapitre concerne l'apport de ces mesures dans le cadre de protocoles de reconnaissance de visage en mode dégradé. Deux types de scénarios ont été étudiés. Le premier consiste à comparer un modèle de référence de type "passeport" à un modèle de test de type inconnu et à utiliser la mesure de qualité des images afin de déterminer le choix du pré-traitement et de l'espace de réduction à utiliser. Cette approche a montré son efficacité dans le Test5 où un gain de performance est constaté, en se basant sur cette stratégie et sur une connaissance *a priori* de la qualité. Dans le Test6, nous avons utilisé la même stratégie, mais en se basant sur une classification par les mesures de qualité présentées. Nous avons constaté que cette stratégie de reconnaissance combinée à cet étiquetage automatique des images permet de s'affranchir efficacement d'un étiquetage manuel sans aucune dégradation des performances. Il est ainsi envisageable d'utiliser cette stratégie dans des applications réelles pour lesquelles aucune information *a priori* de la qualité des images n'existe.

Le deuxième type de scénario concerne la reconnaissance du visage en utilisant images de référence de qualité dégradée et des images de test de qualité diverse. Pour les différentes études de ce scénario, nous avons utilisé le même espace de réduction P1 et un pré-traitement basé sur le lissage anisotropique. Les études présentent diverses stratégies de fusion des mesures de qualité et des scores de reconnaissance.

La stratégie la plus efficace pour ce type de scénario est finalement l'utilisation de plusieurs images pour les modèles de référence et de test ainsi que la pondération des scores de comparaison par la qualité de images de test (Test10-D3).

Chapitre 8

Conclusion et Perspectives

Dans cette thèse, nous avons présenté un certain nombre de méthodes et d'approches permettant d'améliorer la reconnaissance du visage en 2D sous des conditions d'illumination et de qualité variables.

Nous avons commencé dans le chapitre 2 par exposer un état de l'art non exhaustif des méthodes de reconnaissance de visage 2D. Les outils d'évaluation biométrique ainsi que les bases de données publiques de visages, utilisés lors de nos évaluations, sont présentés dans le chapitre 3. Une étude comparative des différents algorithmes de reconnaissance (ACP, LDA et DLDA), qui sont basés sur l'approche globale par réduction d'espace, a été présentée dans le chapitre 4. Notre choix s'est finalement porté sur l'algorithme DLDA qui a démontré de meilleures performances que l'ACP et la LDA.

Le chapitre 5 fut consacré à l'étude du comportement des algorithmes de reconnaissance du visage suivant les paramètres de normalisations géométrique et photométrique. Pour la normalisation géométrique, les expériences ont montré que les performances de la reconnaissance se stabilisent à partir d'une taille suffisante des images après normalisation, calculée en terme de distance entre les centres des deux yeux. Au delà d'une distance de 50 pixels entre les centres des yeux, la variation des performances de reconnaissance n'est plus significative. La deuxième partie de ce chapitre est consacrée à la normalisation photométrique. Dans une première phase, nous avons commencé par étudier l'influence de trois méthodes de transformation de l'image couleur en niveaux de gris. La transformation utilisée est rarement citée dans la littérature bien que les résultats de la reconnaissance

soient très sensibles à ce choix. Nous avons ensuite étudié les principales méthodes de correction de l'illumination qui ont été présentées dans la littérature. Dans des conditions d'acquisition dégradées, la méthode basée sur le lissage anisotropique a montré sa supériorité, en terme de performances, par rapport à l'égalisation d'histogramme, à la correction Gamma et à l'approche MultiRetinex. Cette différence de performance n'est plus vérifiée lorsque les images de référence et les images de test sont de bonne qualité (illumination et résolution). A la fin de ce chapitre, nous avons présenté une nouvelle approche de normalisation photométrique qui a montré son efficacité par rapport aux méthodes classiques étudiées dans le cas de tests en conditions dégradées. Cette normalisation photométrique est basée sur la reconstruction de la luminance du visage en exploitant son aspect symétrique. Les résultats des évaluations sur les bases FRGC et BANCA ont montré que dans le cas d'images acquises dans des conditions d'illumination et de mise au point contrôlées, un pré-traitement simple, comme l'égalisation d'histogramme, est largement suffisant. En revanche, une dégradation des performances peut subvenir lorsqu'un pré-traitement non adapté à la qualité des images de référence et des images de test est appliqué. C'est donc dans ce contexte que la nouvelle méthode proposée a toute son importance.

Dans le chapitre 6, nous avons proposé d'utiliser l'information spatio-fréquentielle du visage pour obtenir une bonne représentation. Cette information est extraite de l'image des pixels par application de filtres de Gabor à différentes résolutions et à diverses orientations. Cette représentation a largement été utilisée et étudiée dans la littérature. Dans la majorité de ces études, seule l'amplitude de la réponse des filtres de Gabor a été employée alors que la phase est omise. Nous avons expliqué les raisons de la limitation de l'utilisation de cette phase et nous avons fourni une solution simple pour pallier cette limitation. Nous avons également proposé d'utiliser une nouvelle représentation du visage basée sur la fusion de l'amplitude et de la phase corrigée. Les résultats des différentes évaluations sur les deux bases de données FRGC et BANCA ont confirmé la robustesse de cette fusion. Dans la première campagne d'évaluation de la base IV², cette méthode a montré sa supériorité par rapport à plusieurs algorithmes classiques comme la LDA, la PCA modulaire. L'amélioration est surtout remarquable lorsque les conditions des tests sont dégradées. Nous nous sommes aussi intéressés à l'influence du choix des paramètres des différents filtres de Gabor comme les orientations, les résolutions et les fréquences sur les performances. Nous avons observé que l'algorithme de la DLDA

est capable d'extraire l'information discriminante de cette nouvelle représentation même si une information non discriminante est rajoutée à partir d'une certaine résolution d'analyse.

Nous avons ainsi proposé un système complet comprenant prétraitement et codage global avec filtres de Gabor, présentant d'excellentes performances même en conditions dégradées.

En nous basant sur les observations du chapitre 5, liées à l'apport des différents pré-traitements selon la qualité des images, nous avons proposé, dans le chapitre 7, une stratégie de reconnaissance liée à la qualité des images de référence et de test. Nous avons commencé par étudier différentes mesures afin de pouvoir classer les images selon leur qualité en terme de netteté et d'illumination.

Le premier critère mesure le déséquilibre de l'illumination par calcul des distributions locales des statistiques des niveaux de gris. Plusieurs statistiques locales ont été étudiées et la médiane a donné les meilleures performances de séparation sur la base d'apprentissage.

Un deuxième critère exploite la différence des histogrammes des images de luminance entre les images acquises dans des conditions d'illumination contrôlées et les images acquises dans des conditions d'illumination variables. On mesure alors la distance du vecteur de l'histogramme de l'image luminance à l'hyperplan séparateur calculé par SVM. L'apprentissage de cet hyperplan est effectué sur une partie de la base d'apprentissage de la base de données de FRGC. Ces deux mesures de l'illumination sont intéressantes parce qu'elles sont complémentaires. Enfin, afin de pouvoir mesurer le niveau du flou de l'image du visage, nous avons proposé une méthode basée sur une approche fréquentielle locale pondérée. Cette méthode a donné des résultats meilleurs que deux autres méthodes classiques de calcul du flou, à savoir une méthode basée sur l'approche gradient et une deuxième basée sur l'approche fréquentielle globale. Ceci s'explique par le fait que ce nouveau critère est plus adapté à la nature des images du visage que les deux autres mesures. La classification des images, en terme de qualité, par fusion de ces différents critères est réalisée par la méthode non-supervisée des k-moyennes. Nous avons validé notre système de classification sur la base BANCA.

Un bon accord sur les données considérées comme dégradées est obtenu par notre classificateur et par celui des organisateurs des évaluations. Par contre sur les images considérées comme contrôlées dans les évaluations, nous notons quelques

différences avec notre classification.

En s'appuyant sur cette méthode de classification des images par la qualité, nous avons étudié, par la suite, plusieurs stratégies de reconnaissance. Pour cela, un nouveau découpage de la base FRGC a été nécessaire et plusieurs scénarios de tests ont été construits. Lors de l'utilisation de plusieurs images en test, nous avons obtenu une amélioration relative de 40% dans les performances grâce à l'utilisation d'une information de qualité pour pondérer les scores résultant de chaque test. Nous améliorons ainsi significativement les performances par cette stratégie au regard de la seule utilisation des coefficients de Gabor, sans tenir compte de la qualité.

Les perspectives de ce travail de thèse restent nombreuses. Nous n'avons fait qu'effleurer l'apport qui peut être attendu dans le traitement d'images vidéo (application de vidéo surveillance par exemple) où plusieurs images de la personne sont disponibles.

Dans cette direction, il serait intéressant de considérer d'autres moyens d'introduire la mesure de qualité dans le calcul des scores. Nous avons effectivement proposé une méthode de fusion de score par simple moyenne pondérée des scores où la mesure de la qualité sert de poids. Cependant, d'autres méthodes utilisant des modèles statistiques comme les GMM ou les SVM seraient intéressantes à étudier.

Un autre point à explorer est l'utilisation d'une modélisation synthétique du visage afin de caractériser un visage de bonne ou de mauvaise qualité. Nécessitant une phase d'apprentissage, le point faible de notre approche reste l'étiquetage manuel de la base d'apprentissage. Cet étiquetage est assez subjectif et sujet à des erreurs dues à l'appréciation des opérateurs. L'utilisation d'un modèle de synthèse génératif d'un visage de bonne ou de mauvaise qualité, permettrait sans doute de diminuer, et voir éliminer, l'influence des erreurs d'étiquetage qui peuvent entacher notre méthode.

Les dernières avancées de la technologie de capture 3D des images de visage ont permis de mettre en place des systèmes de reconnaissance assez robustes par rapport à la pose, s'appuyant sur les informations supplémentaires que le modèle 3D peut fournir pour l'identification. En principe, la modélisation 3D d'un visage est considérée comme robuste, voir même indépendante des problèmes environne-

mentaux d'acquisition. En pratique, la capture d'un modèle précis du visage 3D est sujette toutefois à des variations dans les conditions d'illumination. Il serait intéressant d'étudier, dans ce cas, l'influence de l'illumination sur les performances et d'introduire de nouvelles mesures qui correspondent aux problèmes que rencontre l'acquisition 3D, tels que le pourcentage d'occultation des parties du visage et les erreurs de reconstruction du modèle.

Une autre piste pour pallier aux problèmes de variabilité de l'illumination est l'utilisation des images en proche infrarouge. Une étude avec ce type d'image a été entamée au cours de ma thèse en collaboration de Walid Hizem et les résultats sont encourageants [96].

Finalement, une voie que nous souhaitons explorer est celle de l'implémentation de notre système complet sur une plate-forme nomade. Une étude préliminaire a été réalisée dans le cadre du projet français OPPIDUM **VINSI** "Vérification d'Identité Numérique Sécurisée Itinérante" coordonné par *Thales*. Pour cela, le choix automatique de la stratégie de reconnaissance ainsi que l'introduction de nos mesures de qualité dans le calcul du score doivent être optimisés. Les problèmes à résoudre sont essentiellement liés à la complexité du système actuel en terme de calcul et de mémoire pour des intégrations futures.

Une partie de ce travail d'optimisation a néanmoins été déjà effectuée afin de pouvoir obtenir des résultats dans des temps convenables pour les protocoles de la grande base de données FRGC.

Liste des publications

Chapitre de livre :

(2008) Massimo Tistarelli, Manuele Bicego, José L. Alba-Castro, Daniel Gonzalez-Jimenez, Mohamed-Anouar Mellakh, Albert Ali Salah, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. **2D Face Recognition**. Dijana Petrovska-Delacrétaz, editor, **Guide to Biometric Reference Systems and Performance Evaluation**, chapter 8, Springer, 2009.

Articles de journaux

(2009) Walid Hizem, Lorène Allano, Anouar Mellakh and Bernadette Dorizzi : **Face recognition from synchronized visible and near infrared images**. IET Signal Processing, 2009.

Conférences internationales :

(2004) Emine Krichen, M. Anouar Mellakh, Sonia Garcia-Salicetti, Kamel Hamrouni, Nouredine Ellouze and Bernadette Dorizzi : **Iris Identification Using Wavelet Packet for Images in Visible Light Illumination**. ICBA 2004 : 491-497

(2004) Emine Krichen, M. Anouar Mellakh, Sonia Garcia-Salicetti and Bernadette Dorizzi : **Iris Identification Using Wavelet Packets**. ICPR (4) 2004 : 335-338

- (2005) S. Garcia-Salicetti, Anouar Mellakh, L. Allano and B. Dorizzi : **A Generic Protocol for Multibiometric Systems Evaluation On Virtual and Real Subjects**. Workshop AVBPA 2005, USA, July 2005.
- (2005) S. Garcia-Salicetti, Anouar Mellakh, L. Allano and B. Dorizzi : **Multimodal biometric score fusion : the mean rule vs. support vector classifiers**.EUSIPCO 2005, Antalya, Turkey.
- (2006) M.A. Mellakh, D. Petrovska-Delacretaz, and B. Dorizzi : **Using signal/residual information of eigenfaces for pca face space dimensionality characteristics**. Pattern Recognition, 2006. ICPR 2006. 18th International Conference on Pattern Recognition, 4 pages :574-577, 2006.
- (2007) M. Anouar Mellakh, Dijana Petrovska-Delacr  az, and Bernadette Dorizzi : **Combinaison de classifieurs pour la reconnaissance du visage**. TAIMA'07 : Traitement et Analyse de l'Information : m  thodes et applications, 22-26 mai , Hammamet, Tunisie, 2007
- (2008) D. Petrovska Delacretaz, S. Lelandais, J. Colineau, L. Chen, B. Dorizzi, M. Ardabilian, E. Krichen, M.A. Mellakh, A. Chaari, S. Guer, J. D Hose, and B.B. Amor : **The IV2 multimodal biometric database (including iris, 2d, 3d, stereoscopic, and talking face data)**, and the iv2-2007 evaluation campaign. In BTAS08, pages 1-7, 2008.

Bibliographie

- [1] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. *Face Recognition with Local Binary Patterns*. 2004.
- [2] O. Ayinde and Y.H. Yang. Face recognition approach based on rank correlation of gabor-filtered images. *Pattern Recognition*, 35(6) :1275–1289, June 2002.
- [3] FR Bach and MI Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, pages 1–48, 2002.
- [4] E. Bailly-Baillière, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariéthoz, J. Matas, K. Messer, V. Popovici, F. Porée, B. Ruiz, and J.-P. Thiran. The BANCA Database and Evaluation Protocol. In *4th International Conference on Audio-and Video-Based Biometric Person Authentication (AVBPA'03), volume 2688 of Lecture Notes in Computer Science*, pages 625–638, Guildford, UK, January 2003. Springer.
- [5] H.G. Barrow and J.M. Tenenbaum. Recovering intrinsic scene characteristics from images. In *CVS78*, pages 3–26, 1978.
- [6] A. Bartlett and JR Movellan. Face recognition by independent component analysis. *TRANSACTIONS ON NEURAL NETWORKS*, 13 :303–321, November 2002.
- [7] P.N. Belhumeur and D.J. Kriegman. What is the set of images of an object under all possible lighting conditions? *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on*, pages 270–277, Jun 1996.
- [8] Alphonse Bertillon. *Ethnographie moderne : les races sauvages*. 1883.

- [9] Christopher M. Bishop. *Information Theory, Inference, and Learning Algorithms*. Springer, 2006.
- [10] R. M. Bolle, N. K. Ratha, and S. Pankanti. Error analysis of pattern recognition systems-the subsets bootstrap. *Computer Vision and Image Understanding*, 93(1) :1–33, January 2004.
- [11] David S. Bolme, J. Ross Beveridge, Marcio Teixeira, and Bruce A. Draper. The csu face identification evaluation system : Its purpose, features, and structure. In *ICVS*, pages 304–313, 2003.
- [12] Michael Brauckmann and Martin Werner. Quality values for face recognition. Technical report, VIISAGE, 2006.
- [13] Michael Bulmer. *Francis Galton : pioneer of heredity and biometry*, volume 1. Baltimore, Md. London , Johns Hopkins University Press – cop, 2003.
- [14] Kyong I. Chang, Kevin W. Bowyer, and Patrick J. Flynn. Face recognition using 2d and 3d facial data. *Workshop in Multimodal User Authentication*, pages 25–32, 2003.
- [15] Kyong I. Chang, Kevin W. Bowyer, and Patrick J. Flynn. An evaluation of multimodal 2d+3d face biometrics. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(4) :619–624, 2005.
- [16] L.F. Chen, H.Y.M. Liao, M.T. Ko, J.C. Lin, and G.J. Yu. A new lda-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 33(10) :1713–1726, October 2000.
- [17] M. N. Dailey, W. C. Cottrell, C. Padgett, and R. Adolphs. A neural network that categorizes facial expressions. *Journal of Cognitive Science*, pages 1158–1173, 2002.
- [18] J. Daugman. How iris recognition works. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1) :21–30, Jan. 2004.
- [19] R.L. de Queiroz and K.M. Braun. Color embedding into gray images. pages III : 944–947, 2005.
- [20] B. Fasel. Multiscale facial expression recognition using convolutional neural networks. in *Proc. of the third Indian Conference on Computer Vision (ICVGIP)*, 2002.

- [21] R.A Fisher. The use of multiple measures in taxonomic problems. *Ann. Eugenics*, 7 :179–188, 1936.
- [22] X.F. Gao, S.Z. Li, R. Liu, and P.R. Zhang. Standardization of face image sample quality. In *ICB07*, pages 242–251, 2007.
- [23] A. S. Georgiades, D. J. Kriegman, and P. N. Belhumeur. Illumination cones for recognition under variable lighting : Faces. In *CVPR '98 : Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, page 52, Washington, DC, USA, 1998. IEEE Computer Society.
- [24] A.S. Georgiades, P.N. Belhumeur, and D.J. Kriegman. From few to many : illumination cone models for face recognition under variable lighting and pose. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(6) :643–660, Jun 2001.
- [25] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing (2nd Edition)*. Prentice Hall, January 2002.
- [26] Ralph Gross and Vladimir Brajovic. An image preprocessing algorithm for illumination invariant face recognition. In *4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*. Springer, June 2003.
- [27] Abdenour Hadid. Face description with local binary patterns : Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(12) :2037–2041, 2006.
- [28] Ziad M. Hafed and Martin D. Levine. Face recognition using the discrete cosine transform. *Int. J. Comput. Vision*, 43(3) :167–188, 2001.
- [29] Wonjun Hwang, Gyutae Park, Jongha Lee, and Seok-Cheol Kee. Multiple face model of hybrid fourier feature for large face image set. In *CVPR '06 : Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1574–1581, Washington, DC, USA, 2006. IEEE Computer Society.
- [30] Buciu I., kotropoulos C., and Pitas I. Ica and gabor representation for facial expression recognition. *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, 2 :II–855–8 vol.3, 14-17 Sept. 2003.

- [31] Phillips J., H. Moon, S. Rizvi, and P. Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22, 2000.
- [32] Phillips P. J., Wechsler H., Huang J., and Rauss P. The feret database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 1998.
- [33] Paul T. Jackway and Mohamed Deriche. Scale-space properties of the multiscale morphological dilation-erosion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1) :38–51, 1996.
- [34] Anil K. Jain and Farshid Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recogn.*, 24(12) :1167–1186, 1991.
- [35] Anil K. Jain and Stan Z. Li. *Handbook of Face Recognition*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
- [36] Phillips Jonathon, P. Grother, and ROSS J. Face recognition vendor test 2002. Technical report, NIST, 2003.
- [37] Phillips Jonathon, H. Wechsler, J. Huang, and P. Rauss. The feret database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16 :295–306, 1998.
- [38] Michael David Kelly. *Visual identification of people by computer*. PhD thesis, Stanford, CA, USA, 1971.
- [39] Kirby and M Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. *Pattern Analysis and Machine Intelligence*, 12 :103–108, Jan 1990.
- [40] C. Kotropoulos, A. Tefas, and I. Pitas. Morphological elastic graph matching applied to frontal face authentication under optimal and real conditions. *Multimedia Computing and Systems, 1999. IEEE International Conference on*, 2 :934–938 vol.2, Jul 1999.
- [41] C.L. Kotropoulos, A. Tefas, and I. Pitas. Frontal face authentication using discriminating grids with morphological feature vectors. *Multimedia Computing and Systems, IEEE Transactions on*, 2(1) :14–26, Mar 2000.

- [42] Norbert Krüger. An algorithm for the learning of weights in discrimination functions using a priori constraints. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7) :764–768, 1997.
- [43] K. Kryszczuk and A. Drygajlo. On face image quality measures. *Multimodal User Authentication*, May 2006.
- [44] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Wurtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *Transactions on Computers*, 42(3) :300–311, Mar 1993.
- [45] C J Lee and S D Wang. Fingerprint feature extraction using gabor filters. *Electronics Letters*, 1999.
- [46] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755) :788–791, October 1999.
- [47] Ying li Tian, Takeo Kanade, and Jeffrey F. Cohn. Evaluation of gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. In *FGR '02 : Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, page 229, Washington, DC, USA, 2002. IEEE Computer Society.
- [48] Chengjun Liu. Gabor-based kernel pca with fractional power polynomial models for face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(5) :572–581, 2004.
- [49] Chengjun Liu. Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. *Pattern Analysis and Machine Intelligence*, 28 :725–737, May 2006.
- [50] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition. *PRL*, 26(2) :181–191, January 2005.
- [51] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki. The det curve in assessment of detection task performance. In *Eurospeech*, pages 1895–1898, 1997.
- [52] AM Martinez and AC Kak. Pca versus lda. *Pattern Analysis and Machine Intelligence*, 23 :228–233, 2001.

- [53] Aurélien Mayoue and Dijana Petrovska-Delacrètaz. Open source reference systems for biometric verification of identity. *The 4th International Conference on Open Source Systems*, September 2008.
- [54] M.A. Mellakh, D. Petrovska-Delacretaz, and B. Dorizzi. Using signal/residual information of eigenfaces for pca face space dimensionality characteristics. *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, 4 :574–577, 0-0 2006.
- [55] Kieron Messer and al. Face authentication test on the banca database. In *ICPR '04 : Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) volume 4*, pages 523–532, Washington, DC, USA, 2004. IEEE Computer Society.
- [56] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.R. Mullers. Fisher discriminant analysis with kernels. *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*, pages 41–48, Aug 1999.
- [57] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, A. Smola, and K. Müller. Invariant feature extraction and classification in kernel spaces, 2000.
- [58] H. Moon and P.J. Phillips. Analysis of pca-based face recognition algorithms. *Empirical Evaluation Techniques in Computer Vision*, 1998.
- [59] NIST. The 2008 nist speaker recognition evaluation, June 2008.
- [60] University of Bologna, Michigan State University, San Jose State University, and Universidad Autonoma de Madrid. the fourth international fingerprint verification competition, June 2006.
- [61] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7) :971–987, 2002.
- [62] A. V. Oppenheim and J. S. Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5) :529–541, May 1981.
- [63] A.V. Oppenheim and J.S. Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5) :529–541, May 1981.
- [64] P Penev and L Sirovich. The global dimensionality of face space. *Face and Gesture*, pages 264–270, 2000.

- [65] P.S. Penev and J.J. Atick. Local feature analysis : A general statistical theory for object representation. *Network : Computation in Neural Systems*, 7(3) :477–500, 1996.
- [66] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, pages 84–91, Jun 1994.
- [67] D. Petrovska Delacretaz, S. Lelandais, J. Colineau, L. Chen, B. Dorizzi, M. Ardabilian, E. Krichen, M.A. Mellakh, A. Chaari, S. Guerfi, J. D'Hose, and B.B. Amor. The iv2 multimodal biometric database (including iris, 2d, 3d, stereoscopic, and talking face data), and the iv2-2007 evaluation campaign. In *BTAS08*, pages 1–7, 2008.
- [68] Dijana Petrovska-Delacrètaz, Gérard Chollet, and Bernadette Dorizzi. *Guide to Biometric Reference Systems and Performance Evaluation*. Springer, 2009.
- [69] Jonathon Phillips and Patrick J Flynn. Overview of the face recognition grand challenge. *Computer Vision and Pattern Recognition*, june 2005.
- [70] P. Jonathon Phillips, W. Todd Scruggs, Alice J. O Toole, Patrick J. Flynn, Kevin W. Bowyer, Cathy L. Schott⁵, and Matthew Sharpe. Fvrt 2006 and ice 2006 large-scale results evaluation report. Technical, National Institute of Standards and Technology, March 2007.
- [71] S. Der Phillips J., P. J. Rauss. Feret recognition algorithm development and test results. *Army Research Laboratory technical report, ARL-TR-995*, 1996.
- [72] Norman Poh, Guillaume Heusch, and Josef Kittler. On combination of face authentication experts by a mixture of quality dependent fusion classifiers. In *MCS*, pages 344–356, 2007.
- [73] Laiyun Qing, Shiguang Shan, and Xilin Chen. Face relighting for face recognition under generic illumination. *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, 5 :V–733–6 vol.5, 17-21 May 2004.

- [74] Laiyun Qing, Shiguang Shan, and Wen Gao. Face recognition under varying lighting based on derivatives of log image. In *SINOBIOMETRICS*, pages 196–204, 2004.
- [75] David G. Stork Richard O. Duda, Peter E. Hart. *Pattern Classification*. Wiley-Interscience, 2001.
- [76] Tammy Riklin-Raviv and Amnon Shashua. The quotient image :class based recognition and synthesis under varying illumination conditions. *cvpr*, 02 :2566, 1999.
- [77] Daniel L. Ruderman. The statistics of natural images. *Network : Computation in Neural Systems*, pages 517–548., 1994.
- [78] F.S. Samaria and S. Young. Hmm-based architecture for face identification. *IVC*, 12(8) :537–543, October 1994.
- [79] M. Savvides, R. Abiantun, J. Heo, S. Park, C. Xie, and B.V.K. Vijaya Kumar. Partial and holistic face recognition on frgc-ii data using support vector machine. In *Biometrics06*, page 48, 2006.
- [80] M. Savvides and B.V.K. Vijaya Kumar. Quad phase minimum average correlation energy filters for reduced memory illumination tolerant face authentication. In *AVBPA03*, pages 19–26, 2003.
- [81] B Scholkopf, A Smola, and KR Muller. Nonlinear component analysis as a kernel eigenvalue problem. *Technical Report No 44*, December 1996.
- [82] M. Schulze, K. Scheffler, and K. W. Omlin. Recognizing facial actions with support vector machines. in *Proc. PRASA*, pages 93–96, 2002.
- [83] A. Shashua and T. Riklin-Raviv. The quotient image : class-based re-rendering and recognition with varying illuminations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(2) :129–139, Feb 2001.
- [84] L.L. Shen and L. Bai. Gabor feature based face recognition using kernel methods. In *AFGR04*, pages 170–175, 2004.
- [85] Terence Sim, Simon Baker, and Maan Bsat. The cmu pose, illumination, and expression (pie) database, 2002.
- [86] Terence Sim and Takeo Kanade. Illuminating the face. Technical Report CMU-RI-TR-01-31, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, September 2001.

- [87] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A*, 4(3) :519, 1987.
- [88] Y. Su, S.G. Shan, X.L. Chen, and W. Gao. Hierarchical ensemble of global and local classifiers for face recognition. In *ICCV07*, pages 1–8, 2007.
- [89] Daniel L. Swets and John (Juyang) Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8) :831–836, 1996.
- [90] L. Szirmay-Kalos. Monte-carlo methods in global illumination, 1999.
- [91] Anastasios Tefas, Constantine Kotropoulos, and Ioannis Pitas. Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(7) :735–746, 2001.
- [92] Anastasios Tefas, Constantine Kotropoulos, and Ioannis Pitas. Face verification using elastic graph matching based on morphological signal decomposition. *Signal Process.*, 82(6) :833–851, 2002.
- [93] M. Turk and A. Pentland. Eigenfaces for recognition. *journal of Cognitive Neuroscience*, 3 :71–86, 1991.
- [94] Zia ur Rahman, Glenn A Woodell, and Daniel J Jobson. A comparison of the multiscale retinex with other image enhancement techniques. In *Proceedings of the IST 50th Annual Conference*, 1997.
- [95] B. Draper W. Yambor and R. Beveridge. Analyzing pca-based face recognition algorithms : Eigenvector selection and distance measures. *Empirical Evaluation Methods in Computer Vision*, World Scientific Press, 2002.
- [96] Anouar Mellakh Walid Hizem, Lorène Allano and Bernadette Dorizzi. Face recognition from synchronized visible and near infrared images. *IET Signal Processing*, 2009.
- [97] Haitao Wang, Stan Z. Li, Yangsheng Wang, and Weiwei Zhang. Illumination modeling and normalization for face recognition. In *AMFG '03 : Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, page 104, Washington, DC, USA, 2003. IEEE Computer Society.
- [98] Frank Weber. Some quality measures for face images and their relationship to recognition performance. Technical report, Cognitec, 2006.

- [99] L. Wiskott. Phantom faces for face analysis. In *ICIP '97 : Proceedings of the 1997 International Conference on Image Processing (ICIP '97) 3-Volume Set-Volume 3*, page 308, Washington, DC, USA, 1997. IEEE Computer Society.
- [100] Laurenz Wiskott, Jean-Marc Fellous, Norbert Kruger, and Christoph von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7) :775–779, 1997.
- [101] Laurenz Wiskott and Christoph von der Malsburg. Recognizing faces by dynamic link matching. In Axel Wismüller and Dominik R. Dersch, editors, *Symposion über biologische Informationsverarbeitung und Neuronale Netze - SINN '95*, pages 63–68, München, 1996.
- [102] Rolf P. Würtz. Object recognition robust under translations, deformations, and changes in background. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7) :769–775, 1997.
- [103] Hua Yu and Jie Yang. A direct lda algorithm for high-dimensional data - with application to face recognition. *Pattern Recognition*, 34(10) :2067–2070, 2001.
- [104] J. Zhang, Y. Yan, and M. Lades. Face recognition : Eigenface, elastic matching, and neural nets. *PIEEE*, 85(9) :1423–1435, September 1997.
- [105] Mian Zhou and Hong Wei. Face verification using gaborwavelets and ada-boost. In *ICPR '06 : Proceedings of the 18th International Conference on Pattern Recognition*, pages 404–407, Washington, DC, USA, 2006. IEEE Computer Society.

Remerciements

Certaines personnes trouveront peut être inadéquat que mes remerciements soient mis à la fin de ce manuscrit. Pour moi, c'est leur place idéale, ils s'inscrivent dans la continuité de cette thèse. Après des années de travail, je trouve que finir ces pages en remerciant les personnes qui méritent de l'être est plus valorisant et sincère.

Pour ne pas déroger à la tradition et en toute sincérité, je tiens à remercier les membres du jury qui ont accepté d'évaluer mon travail. Juger un travail est loin d'être une tâche des plus faciles. Je tiens donc à remercier mon rapporteur le Professeur Sylvie Lelandais pour la précision, pour la pertinence de ses critiques et pour ses conseils qui ont permis d'améliorer la compréhension de ce manuscrit. Mes remerciements s'adressent également au Professeur Faouzi Gorbél pour ses critiques très constructives et les idées très intéressantes qu'il a formulées dans son rapport et lors de la présentation. Je remercie au même titre le Professeur André Gagalowicz d'avoir accepté de faire parti du jury pour juger mon travail et enfin, un grand merci au Professeur Maurice Milgram d'avoir accepté de présider le jury et pour ses critiques lors de la soutenance.

Je tiens particulièrement à remercier ma directrice de thèse, le Professeur Bernadette Dorizzi, avec qui j'ai eu le plaisir, la chance et l'honneur de travailler, de discuter et d'avancer dans ma thèse. Je tiens sincèrement à la remercier pour son soutien, non seulement, sur le plan scientifique mais aussi sur le plan humain. Mes remerciements s'adressent aussi à mon encadreur de thèse, le Docteur Dijana Petrovska, qui m'a appris à analyser, à critiquer et à m'exprimer le plus clairement possible : en particulier, à penser et à agir comme un scientifique et non pas comme l'ingénieur débutant que j'étais au commencement de ma thèse.

Je tiens bien sûr à remercier mes parents pour leur patience et leur dévoue-

ment. Je pense que la distance est d'autant plus dure pour eux que pour moi, mais sans leurs encouragements, je ne serais pas entrain d'écrire ces quelques mots. Alors merci maman, merci papa. Je n'oublie pas non plus de remercier mon frère et ma petite soeur pour leur patience.

Mes remerciements s'adressent également à tous mes collègues et amis de l'Institut National des Télécommunications que j'ai eu la chance de côtoyer pendant des années et qui m'ont permis de gérer des moments de stress et de fatigue. Je pense particulièrement à Yann, Murielle, Yanneck, Patricia, Yollande, Aurélien, François et Marie-Thérèse.

J'ai eu la chance de faire la rencontre d'un ami, il y a près d'une dizaine d'année, sur qui j'ai pu compter à tout moment et que je considère comme un frère. Mon ami nEzmine (avec un E, s'il vous plaît) est et restera pour moi l'ami de confiance qu'on a intérêt à avoir. Même si je ne lui confiais pas mes clés. Merci Emine et au plaisir de re-travailler avec toi.

Lors des Doctoriales 2005, j'ai fais la connaissance de jeunes doctorants de disciplines différentes qui sont devenus des amis très chers et des complices d'aventures. Je remercie Christina, Lorène, Isabelle, Delphine, Ons, Coralie, Walid, Etienne, Xavier, Charles, Benoit, et Benjamin. Chacun a fait son petit chemin et nos longues soirées ou dîners autour d'excellents plats resteront parmi les plus merveilleux souvenirs de mes dernières années.

Je tiens à saluer la mémoire de ma seconde mère Christine Villiers, qui a été pour moi une vraie mère de substitution et qui m'a poussé avec toute sa générosité, sa force et son courage à aller de l'avant. Je la remercie à titre posthume d'avoir été pour moi une amie et une complice. Nos petites discussions me manquent et ton souvenir ne s'éteindra jamais. Je tiens à remercier par la même occasion Alain Amesland pour avoir été présent tout au long de ces moments.

Enfin, mes remerciements s'adressent à la femme de ma vie et ma complice Tiphaine pour sa patience, pour sa présence, pour son dévouement et pour son aide inestimable dans l'accomplissement de ce travail. Ses encouragements et son dévouement m'ont permis dans les moments les plus difficiles à repousser mes limites. Je la remercie pour sa présence ces dernières années et je lui promets, même si ce n'est peut-être pas l'endroit, de faire de mon mieux pour la rendre heureuse.

Ces remerciements peuvent vous sembler un peu longs, mais quelques lignes ne pourront pas résumer des années de rencontres, d'amitiés et d'aventures.

TITRE**Reconnaissance des visages en conditions dégradées**

RÉSUMÉ

De nos jours, les algorithmes de reconnaissance du visage proposés dans la littérature ont atteint un niveau de performance correct lorsque les conditions d'acquisition des images comparées sont contrôlées, mais les performances chutent lorsque ces conditions sont dégradées. Les conditions d'acquisition contrôlées correspondent à un bon équilibre d'illumination, ainsi qu'une haute résolution et une netteté maximale de l'image du visage.

Afin de bien cerner le problème de dégradation des performances dans les conditions d'acquisition difficiles et de proposer par la suite des solutions adaptées, nous avons effectué plusieurs études à différents niveaux de la chaîne de la reconnaissance. Ces études concernent le comportement des algorithmes basés sur les approches globales. Elles concernent également les différentes méthodes de normalisation photométrique ainsi que des stratégies de reconnaissance basées sur la qualité des images du visage.

Les solutions proposées à chaque niveau de cette chaîne ont apporté une amélioration significative des performances par rapport aux approches classiques. Pour les algorithmes de reconnaissance, nous avons proposé l'utilisation de la fusion de la phase et de l'amplitude des représentations de Gabor de l'image du visage comme nouveau gabarit, à la place de l'image des pixels. Bien que les représentations de Gabor aient été largement utilisées, particulièrement dans les algorithmes basés sur les approches globales, la phase n'a jamais été exploitée jusqu'à ce jour. Nous expliquons dans cette thèse les problèmes liés à l'utilisation de cette phase de Gabor et nous proposons une solution pour y remédier.

Différentes méthodes de normalisation photométrique de l'image du visage ont été étudiées et comparées. Nous avons, par la suite, proposé une nouvelle approche de normalisation basée sur la correction de la composante luminance. Enfin, nous avons présenté une stratégie de reconnaissance se basant sur la mesure de qualité du visage. Cette mesure est une fusion de plusieurs critères de qualité et selon nos expériences, cette stratégie permet une amélioration notable des performances par rapport aux méthodes classiques.

Les différentes études, la validation de nos mesures de qualité ainsi que la validation de la stratégie de reconnaissance ont été effectuées sur les deux bases de données publiques de visage FRGCv2 et BANCA largement utilisées par la communauté.

MOT-CLEFS

Biométrie, Reconnaissance de visage, Illumination, Normalisation photométrique, filtres de Gabor, Qualité.

ISBN : 00000000000