



HAL
open science

Reconnaissance Biométrique par Fusion Multimodale du Visage 2D et 3D

Abdelmalik Ouamane

► **To cite this version:**

Abdelmalik Ouamane. Reconnaissance Biométrique par Fusion Multimodale du Visage 2D et 3D. Vision par ordinateur et reconnaissance de formes [cs.CV]. University Mohamed Khider Biskra, Algeria, 2015. Français. NNT: . tel-01852447

HAL Id: tel-01852447

<https://theses.hal.science/tel-01852447>

Submitted on 1 Aug 2018

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

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

Université Mohamed Khider – Biskra

Faculté des Sciences et de la

Technologie

Département: Génie Electrique

Ref :



جامعة محمد خيضر بسكرة

كلية العلوم و التكنولوجيا

قسم: الهندسة الكهربائية

المرجع:

Thèse présentée en vue de l'obtention
Du diplôme de
Doctorat en sciences en Electronique

Intitulé
**Reconnaissance Biométrique par Fusion Multimodale
du Visage 2D et 3D**

Présentée par :

Abdelmalik OUAMANE

Soutenue publiquement le 11 Juin 2015

Devant le jury composé de:

SBAA Salim	Maître de Conférences	Univ. Biskra	Président
BENAKCHA Abdelhamid	Maître de Conférences	Univ. Biskra	Rapporteur
BENATIA Mébarka	Maître de Conférences	Univ. Biskra	Co-Rapporteur
TALEB - AHMED Abdelmalik	Professeur	Univ. Valenciennes	Examineur
BENOUDJIT Nabil	Professeur	Univ. Batna	Examineur
BENYOUCEF Moussa	Professeur	Univ. Batna	Examineur
MOUSSAOUI Abdelouahab	Professeur	Univ. Sétif	Examineur

Remerciements

Terminer une thèse de doctorat est comme un marathon. Je n'aurais pas été en mesure de terminer ce marathon sans le soutien d'innombrables personnes au cours des dernières trois années et demi. Je tiens à exprimer mes remerciements et ma vive gratitude à:

- Dr. **BENAKCHA** Abdelhamid, maître de conférences à l'Université Mohammed Kheider – Biskra- directeur de cette thèse. C'est grâce à nos discussions qu'une grande partie de nos travaux a abouti. Il a su faire preuve d'une pédagogie et d'une gentillesse remarquables et j'espère que nous pourrions continuer nos travaux dans le futur. Je le remercie pour ses grandes qualités humaines,

- Dr. **BENATIA BELAHCENE** Mébarka, maître de conférences à l'Université Mohammed Kheider – Biskra-, co-directrice de cette thèse, qui m'a proposé ce sujet et sans laquelle je ne serais sûrement pas arrivé là. Certaines idées sont nées de conversations que nous avons eues ensemble. Je la remercie également pour son temps et son investissement dans tous les aspects de mon travail,

- Pr. **AHMED TALEB Abdelmalik**, Professeur à l'IUT de Valenciennes et Chercheur au LAMIH, pour m'avoir accueilli dans son laboratoire dans des conditions plus que favorables à une recherche fructueuse. Merci Monsieur pour l'aide, le précieux soutien, pour vos encouragements, votre disponibilité et pour l'attention portée à mon travail,

- Dr. **HADID Abdenour**, Chercheur à CMV, Université Oulu – Finlande-, pour l'aide et la correction de nos articles,

- Pr. **BOURENNANE** Salah, Professeur à l'Ecole Centrale de Marseille et responsable du Groupe Signaux Multidimensionnels, pour m'avoir accueilli dans son laboratoire. Nous avons eu d'innombrables échanges scientifiques très constructifs. Ses conseils étaient très bénéfiques et son apport en idées était très important pour nous.

- Dr. **SBAA Salim**, maître de conférences à l'Université Mohammed Kheider – Biskra- qui me fait l'honneur de présider ce jury de thèse,

- Pr. **TALEB – AHMED** Abdelmalik, Pr. **BENOUDJIT** Nabil et Pr. **BENYOUCEF** Moussa de l'Université Hadj Lakhdar – Batna- et Pr. **MOUSSAOUI** Abdelouahab de l'Université Ferhat Abbas – Sétif - qui ont accepté d'être les rapporteurs de ce thèse. Je les remercie tout particulièrement pour l'attention, le temps qu'ils vont consacrer à la lecture de cette thèse. Evidemment, leurs commentaires seront les bienvenus.

Je tiens à remercier tout particulièrement ma famille et mes amis proches:

- **Mon père** qui a toujours su être là dans tous les moments, les bons comme les difficiles ; son soutien et sa générosité constante ont été fondamentaux, sa droiture et sa conduite ont toujours été un modèle pour moi,

- **Ma mère** qui a toujours su m'écouter. Sa douceur, sa tendresse et sa volonté ont toujours mérité mon plus profond respect.

De manière générale, je remercie tous mes amis qui m'ont apporté un grand soutien et permis de passer de très bons moments de joie et de détente qui m'ont été nécessaires.

Je souhaite à présent adresser mes sincères remerciements à toutes les personnes avec qui j'ai eu la chance de travailler ou que j'ai eu l'honneur de côtoyer avant et pendant le déroulement de ma thèse.

Liste des figures	v
Liste des tableaux	vii
Liste des symboles	ix
Résumé	x
ملخص	xi
Introduction générale	1
1. Concepts de base et terminologie en biométrie faciale	4
1.1 Introduction	4
1.1.1 Qu'est-ce que la biométrie?	4
1.1.2 Pourquoi la reconnaissance de visage?	5
1.1.3 Difficultés de reconnaissance de visage 2D	6
1.1.3.1 Illumination	7
1.1.3.2 Pose	7
1.1.3.3 Expressions faciales	
1.1.3.4 Occlusions	8
1.1.3.5 Autres difficultés	7
1.1.4 Bases de données 2D de visage	9
1.1.5 Opportunités de reconnaissance de visage 3D	10
1.2 Reconnaissance de visage 3D	
1.2.1 Techniques d'acquisition de visage 3D	11
1.2.2 Concepts de base et terminologie	13
1.2.3 Base de données 3D de visage	14
1.3 Conclusion	15
2. Etat de l'art en reconnaissance de visages 2D, 3D et 2D+3D	17
2.1 Introduction	17
2.2 Reconnaissance de visage 2D	17
2.2.1 Approches globales	17
2.2.1.1 Les techniques linéaires	18
2.2.1.2 Les techniques non-linéaires	22
2.2.2 Approches locales	23
2.2.2.1 Méthodes locales basées sur les caractéristiques d'intérêt	24
2.2.2.2 Les méthodes locales basées sur l'apparence du visage	25
2.2.3 Autres approches	29
2.2.3. Modèle actif d'apparence (AAM)	30
2.2.4 Discussion	33
2.3 Reconnaissance de visage 3D	33
2.3.1 Approches globales	34
2.3.1.1 Le sous-espace (PCA et LDA)	34
2.3.1.2 Iterative Closest Point (ICP)	34
2.3.1.3 Distance de Hausdorff	36

2.3.1.4	Modèle déformable	37
2.3.2	Approches locales	38
2.3.2.1	Méthodes basées sur le calcul du point	38
2.3.2.2	Méthodes basées sur le calcul de courbe	39
2.3.2.3	Méthodes basées sur les régions	41
2.3.2.4	Les méthodes basées sur les descripteurs	44
2.3.2.5	Les méthodes basées sur la fusion des caractéristiques multiples	45
2.3.3	Autres approches	45
2.3.4	Discussion	47
2.4	Reconnaissance de visage 2D+3D	47
2.5	Conclusion	50
3.	Vérifications uni-modale et multi-algorithmes de visage 2D	51
3.1	Introduction	51
3.2	Vérification uni-modale de visage 2D	51
3.2.1	Prétraitement d'image	51
3.2.1.1	Découpage	51
3.2.1.2	Décimation	52
3.2.1.3	Composante S de l'espace de couleur TSV de l'image	52
3.2.2	Extraction des caractéristiques de visage par les ondelettes de <i>Gabor</i>	53
3.2.3	Algorithmes de réduction d'espace de données	54
3.2.3.1	Analyse en Composantes Principales (ACP)	54
3.2.3.2	Analyse Discriminante Linéaire (LDA)	55
3.2.3.3	Modèle discriminant linéaire amélioré de <i>Fisher</i>	56
3.2.3.4	LDA directe (DLDA)	57
3.2.4	Comparaison	58
3.2.3	Résultats	59
3.2.5.1	Base de données XM2VTS	59
3.2.5.2	Comparaison des algorithmes de réduction de l'espace de données	60
3.2.5.3	Extraction des caractéristiques par les ondelettes de Gabor	61
3.3	Vérification multi-algorithmes de visage 2D	64
3.3.1	Limitations des systèmes biométriques monomodaux	65
3.3.2	Les différentes multimodalités possibles	65
3.3.3	État de l'art de la fusion de scores	67
3.3.3.1	Approche par classification de scores	67
3.3.3.2	Approche par combinaison de scores	67
3.3.4	La fusion de scores	68
3.3.4.1	Normalisation de scores	68
3.3.4.2	Approche par combinaison de scores	70
3.3.4.3	Approche par classification de scores	73
3.3.5	Résultat	76
3.3.5.1	Analyse en Composantes Principales (ACP)	78

3.3.5.2 Comparaison des méthodes de normalisation associées aux méthodes de combinaisons	80
3.3.5.3 Méthodes de normalisation associées aux méthodes de classifications	82
3.4 Conclusion	84
4. Vérification de visage 2D et 3D par CSL et les descripteurs locaux	86
4.1 Introduction	86
4.2 Le système de vérification multimodal proposé	86
4.2.1 Prétraitement	87
4.2.1.1 Alignement	87
4.2.1.2 Image de profondeur et d'intensité	88
4.2.2 Extraction des caractéristiques du visage	90
4.2.2.1 Multi-scale local binary patterns (MLBP)	90
4.2.2.2 Caractéristiques statistiques locales proposées (CSL)	92
4.2.2.3 Ondelettes de Gabor	94
4.2.2.4 Scale-Invariant Feature Transform (SIFT)	94
4.2.3 Scores Fusion	96
4.2.4 Résultats	96
4.2.4.1 Base de données CASIA 3D	96
4.2.4.2 Approche globale (PCA + EFM)	98
4.2.4.3 Multi-échelle LBP (MLBP)	99
4.2.4.4 Caractéristiques statistiques locales (CSL)	99
4.2.4.5 Ondelettes de <i>Gabor</i>	100
4.2.4.6 MLBP + SIFT	101
4.2.4.7 Fusion des méthodes d'extraction de caractéristiques	101
4.3 Extraction de caractéristiques par LPQ et BSIF	101
4.3.1 Extraction de caractéristiques par LPQ (Local Phase Quantization)	101
4.3.1.1 PLQ descripteur de base	102
4.3.1.2 Processus de décorrélation	103
4.3.2 Extraction de caractéristiques par BSIF (Binarized Statistical Image Features)	104
4.3.2.1 Détails	104
4.3.2.2 Mise en œuvre	105
4.3.3 Résultats	107
4.3.3.1 Base de données Bosphorus	107
4.3.3.2 Multi-échelle LBP (MLBP)	108
4.3.3.3 Caractéristiques statistiques locales (CSL)	109
4.3.3.4 Extraction de caractéristiques par le descripteur BSIF (Binarized Statistical Image Features)	110
4.3.3.5 Extraction de caractéristiques par le descripteur LPQ (Local Phase Quantization)	112
4.3.3.6 Fusion des méthodes d'extraction de caractéristiques locales	113
4.3.3.7 Comparaison	114
4.4 Conclusion	115

5. Vérification du visage 2D+3D par statistical LBP et EDA	116
5.1 Introduction	116
5.2 Résumé de notre travail ICIP'2014	116
5.2.1 Expérience dans la base de données XM2VTS	117
5.2.2 Expérience dans la base de données LFW	118
5.3 Description de l'ensemble du système de vérification faciale 2D+3D proposé	119
5.3.1 Prétraitement de visage 2D et 3D	121
5.3.1.1 Détection du bout du nez	121
5.3.2 Extraction des caractéristiques	121
5.3.2.1 Descripteur LBP (Local Binary Patterns)	121
5.3.2.2 Descripteur LPQ (Local Phase Quantization)	124
5.3.2.3 Descripteur BSIF (Binarized Statistical Image Features)	124
5.3.2.4 Descripteur Statistical LBP proposé	126
5.3.3 EDA (Exponential Discriminant Analysis)	128
5.3.3.1 Matrice exponentielle	130
5.3.3.2 EDA	131
5.3.4 WCCN (Within Class Covariance Normalization) et mesure de similarité	133
5.4 Résultats	133
5.4.1 Bases de données	133
5.4.1.1 La base de données FRGC v2	133
5.4.2 Vérification uni-modale de visage 2D et 3D	135
5.4.3 Vérification de visage multi-descripteurs et multimodales (2D + 3D)	147
5.5 Conclusion	151
Conclusion générale et perspectives	153
Publications et Communications	159
Bibliographie	159

1.1 Quelques modalités biométriques (L'image est recueillie à partir d'Internet).	4
1.2 Comparaison de compatibilité MRTD pour diverses caractéristiques biométriques	6
1.3 Exemple d'un visage d'une même personne subissant un changement de luminosité	7
1.4 Exemple d'un visage d'une même personne subissant des variations de pose hors plan	8
1.5 Variabilité intra-classe due à la présence d'expressions faciales	8
1.6 Exemples d'occlusion du visage	9
1.7 Exemple de lumière structurée (à gauche) et sa projection sur un visage	11
1.8 Gauche: modèle de visage capturé en utilisant une configuration de studio de sept caméras, centre: systèmes de capture, droite: modèle de visage capturé avec la caméra stéréo binoculaire des consommateurs	12
1.9 Comparaison de la stéréo photométrie et stéréo photométrie multi spectrale	13
1.10 : Quatre types de formats d'enregistrement du visage 3D, de gauche à droite: texture de visage 3D, image de profondeur, maille en forme de triangle et la zone de nez zoomé, nuages de points et la région de nez zoomée	14
1.11 Balayage de chacune des bases de données du visage les plus utilisées en 3D contenant des variations d'expression	15
2.1 Visages propres (vecteurs propres) des 12 plus grandes valeurs propres sont présentés provenant de la division ATT de la base de données ORL	19
2.2 Certaines images reconstruites sur la base 2-D PCA (en haut) et PCA (en bas) sur la base de données ORL	19
2.3 Vecteurs de base ICA provenant de la division ATT de la base de données ORL	20
2.4 Premiers 12 vecteurs LDA provenant de la division ATT de la base de données ORL	21
2.5 Exemple de grille d'appariement. (a) grille de référence, (b) grille correspondante	25
2.6 Exemple de la méthode ILBP	26
2.7 Exemple de la méthode ELBP	27
2.8 Exemple de la méthode LTP	27
2.9 Exemple de la méthode LBP allongée	28
2.10 Exemple de la méthode MB-LBP	29
2.11 (a) : visage inconnu, (b): initialisation du modèle et (c): adaptation du modèle classique	31
2.12 Classification des principaux algorithmes utilisés en reconnaissance faciale 2D	32
2.13 Exemple 1: recalage entre un modèle 3D et un modèle requête de la même personne avec variations d'expressions faciales	35
2.14 Exemple 2: recalage entre un modèle 3D et un modèle requête de la même personne avec des variations d'expressions faciales plus significatives	35
2.15 Exemple 3: recalage entre un modèle 3D et un modèle requête de deux personnes différentes	35
2.16 Synthèse et transfert de la déformation basés sur la reconnaissance faciale 3D	37
2.17 (a) vecteurs propres 1, 2 et 3, respectivement à partir de la gauche, (b) vecteurs propres 9, 10 et 11 et (c) vecteurs propres 25, 26 et 27	38
2.18 Points-clés détectés sur les scans du visage de cinq personnes différentes pour la combinaison de la base de données FRGC v2 et UND	40
2.19 Comparaisons des courbes radiales et la distance géodésique entre les courbes	41
2.20 Vue d'ensemble de reconnaissance 3D de visage approchée fondée sur l'enregistrement régional	42
2.21 30 régions locales utilisées pour la comparaison du visage	43
2.22 Taux de reconnaissance des régions individuelles sur la base de données FRGC v2	44
2.23 De maille géométrie ver la carte standard	46
2.24 La première ligne: point de vue de rendu solide de ROI, et la deuxième ligne: image de profondeur relative mappée correspondant	46
3. 1 Principe de fonctionnement de notre système de vérification de visage	51
3. 2 Image de visage (a) avant et (b) après découpage 523. 3 Image de visage (a) avant (b) après décimation	52

3. 4	Différents systèmes colorimétriques	52
3. 5	(a) Image découpée et décimée, (b) Image en système TSV, (c) Composante S	53
3. 6	Comparaison entre les projections de deux classes de points ("classe 1" et "classe 2") sur les axes principaux construits par (a) ACP et par (b) LDA	56
3. 7	Exemples des images de la base de données XM2VTS	59
3. 8	Exemples des images d'une personne de la base de données XM2VTS	59
3. 9	Comparaison des TEE pour ACP, LDA, EFM et DLDA dans l'ensemble d'évaluation	60
3. 10	Résultats de la convolution d'une image de visage avec une famille de 40 filtres de Gabor (a) Image en espace de couleur TSV, l'ensemble (b) amplitudes et (c) les phases	61
3. 11	Taux d'égale erreur dans l'ensemble d'évaluation en fonction du seuil T_h	63
3. 12	Les différents systèmes multimodaux	66
3. 13	Normalisation QLQ	70
3. 14	Séparation linéaire dans un espace à deux dimensions	75
3. 15	Distributions de scores des quatre systèmes d'authentification de visages sur la base de données XM2VTS	77
3. 16	Normalisation de scores par la méthode du Min-Max	78
3. 17	Normalisation de scores	79
4. 1	Vue de notre système proposé	87
4. 2	Prétraitement de l'image de profondeur	89
4. 3	Prétraitement de l'image d'intensité	89
4. 4	Opérateur LBP	91
4. 5	(a) Trois voisinages pour des R et P différents, (b) Textures particulières détectées par LBP	91
4. 6	Multi-échelles LBP de l'image de profondeur et d'intensité du visage	92
4. 7	Caractéristiques statistiques locales proposées (CSL)	93
4. 8	Points-clés détectés avec le descripteur SIFT de l'image de profondeur et d'intensité originale et les quatre images <i>LBP</i> associées	95
4. 9	SIFT correspondante entre les visages d'apprentissage et de test appartenant à (a) la même identité et (b) des identités différentes.	95
4. 10	Schéma de la fusion des scores	96
4. 11	Capture des scènes de la base de données CASIA visages 3D	97
4. 12	Variations d'éclairage de la base de données CASIA visage 3D	97
4. 13	Variations d'expressions de la base de données CASIA visage 3D	97
4. 14	Variations des poses de la base de données CASIA visage 3D	98
4. 15	Organigramme de l'ensemble des étapes nécessaire à la construction du descripteur LPQ.	103
4. 16	LPQ, (a) image de profondeur, (b) image d'intensité	104
4. 17	13 images naturelles utilisées pour la formation des Filtres W_i	106
4. 18	Filtres tirés de taille $l=7$ et nombre de bits $n=8$	106
4. 19	Le descripteur BSIF (Binarized Statistical Image)	107
4. 20	Echantillons avec et sans texture (Base de données du Bosphorus)	108
4. 21	Performance de la méthode d'extraction de caractéristiques BSIF (base de données CASIA visage 3D)	110
4. 22	Performance de la méthode d'extraction de caractéristiques BSIF (base de données Bosphorus)	111
5. 1	Schéma général de notre proposition ICIP'14	116
5. 2	Exemples d'images de la base de données LFW	118
5. 3	Exemples d'images de la même personne de la base de données LFW	118
5. 4	Exemples des mêmes et différentes paires de visages dans la base de données LFW	119
5. 5	Schéma général de notre système proposé	120
5. 6	Prétraitement de l'image de profondeur	122
5. 7	Prétraitement de l'image 2D	122

5. 8 Images de visage 2D: (a) avant et (b) après normalisation photométrique par PS	122
5. 9 LBP uniforme $P=8$	123
5. 10 LBP (Local Binary Patterns) (a) image 3D, (b) image 2D	124
5. 11 Local Phase Quantization (LPQ): (a) 3D image, (b) 2D image	124
5. 12 Filtrés construits de taille 9×9 , (a) pour image 3D et (b) pour image 2D	125
5. 13 Descripteur BSIF (Binarized Statistical Image Features)(a) image 3D, (b)image 2D	126
5. 14 Descripteur Statistical LBP (Statistical Local Binary Patterns) $P_1=8, P_2=8$.	128
5. 15 <i>Statistical LBP</i> pour les images de visage 3D	129
5. 16 <i>Statistical LBP</i> pour les images de visage 2D	130
5. 17 Echantillons d'images capturées avec Minolta Vivid 910	134
5. 18 Exemples d'images. (a) Image neutre. (b) Petite expression. (c) Grande expression	135
5. 19 Exemples d'images contenant des artefacts dans l'ensemble FRGC v2	135
5. 20 Courbes ROC pour le descripteur LBP	139
5. 21 Courbes ROC pour le descripteur LPQ	140
5. 22 Courbes ROC pour le descripteur BSIFo	141
5. 23 Courbes ROC pour le descripteur <i>mean LBP</i>	142
5. 24 Courbes ROC pour le descripteur <i>median LBP</i>	143
5. 25 Courbes ROC pour le descripteur <i>var LBP</i>	144
5. 26 Courbes ROC pour le descripteur <i>Skewness LBP</i>	145
5. 27 Courbes ROC pour le descripteur <i>Kurtosis LBP</i>	146
5. 28 ROC III pour la vérification multi-descripteurs	149
5. 29 Vérification multi-descripteurs pour le protocole "neutre contre tout"	150

1.1 Base de données 2D	10
1.2 Quelques bases de données 3D de visage	15
2.1 Liste des variantes récentes de la LBP	29
3. 1 Répartition des photos dans les différents ensembles (LP I)	60
3. 2 Performances des méthodes ACP, LDA, EFM, DLDA	61
1.1 Base de données 2D	10
1.2 Quelques bases de données 3D de visage	15
2.1 Liste des variantes récentes de la LBP	29
3. 1 Répartition des photos dans les différents ensembles (LP I)	60
3. 2 Performances des méthodes ACP, LDA, EFM, DLDA	61
3. 3 Taux d'égale erreur TEE dans l'ensemble d'évaluation pour différents niveaux de résolutions et d'orientations pour les 40 filtres	62
3. 4 Taux d'égale erreur dans l'ensemble d'évaluation pour différentes résolutions et orientations des phases pour les 40 filtres	64
3. 5 Résultats du système d'authentification de visages pour les quatre meilleurs systèmes retenus dans l'ensemble d'évaluation et de test	64
3. 6 Performances des méthodes de normalisation qui ne modifient pas la forme des distributions associées aux méthodes de combinaison de scores simples	80
3. 7 Performances des méthodes de normalisation qui modifient la forme des distributions associées aux méthodes de combinaison de scores simples	81
3. 8 Performances des méthodes de normalisation par logique floue	81
3. 9 Performances des méthodes de normalisation par fusion statistique de Fisher dans l'ensemble d'évaluation et de test	82
3. 10 Performances des méthodes de normalisation associées à SVM dans d'ensemble d'évaluation et de test	83
3. 11 Performances des méthodes de normalisation associées à la classification MLP	83
4. 1 Répartition des photos dans différents ensembles	98
4. 2 Performance de PCA + EFM dans l'ensemble d'évaluation et de test	99
4. 3 Performance de la méthode d'extraction de caractéristiques MLBP dans l'ensemble d'évaluation et de test	99
4. 4 Performance de la méthode d'extraction de caractéristiques CSL dans l'ensemble d'évaluation et de test	100
4. 5 Performance de la méthode d'extraction de caractéristiques par les ondelettes de Gabor dans l'ensemble d'évaluation et de test	101
4. 6 Performance de la méthode d'extraction de caractéristiques MLBP + SIFT dans l'ensemble d'évaluation et de test	101
4. 7 Performance de cinq méthodes d'extraction de caractéristiques dans l'ensemble d'évaluation et de test	102
4. 8 Répartition des photos dans différents ensembles (base de données Bosphorus)	108
4. 9 Performance de la méthode d'extraction de caractéristiques MLBP dans l'ensemble d'évaluation et de test (base de données Bosphorus)	109
4. 10 Performance de la méthode d'extraction de caractéristiques CSL dans l'ensemble d'évaluation et de test (Base de données Bosphorus)	109
4. 11 Performances de la méthode d'extraction de caractéristiques LPQ dans l'ensemble d'évaluation et de test (base de données CASIA visage 3D)	112
4. 12 Performances de la méthode d'extraction de caractéristiques LPQ dans l'ensemble d'évaluation et de test (base de données Bosphorus)	112
4. 13 Performances de quatre descripteurs et leur fusion dans l'ensemble d'évaluation et test (base de données CASIA visage 3D)	113

4. 14 Performances de quatre descripteurs et leur fusion dans l'ensemble d'évaluation et test (base de données Bosphorus)	114
4. 15 Comparaison avec d'autres méthodes locales mentionnées dans la littérature.	115
5. 1 Taux d'erreur total sur l'ensemble d'évaluation et de test (base de données XM2VTS, Protocole de Lausanne I)	117
5. 2 (Mean \pm Standard Error Score) de nos systèmes sur la base de données LFW «Unrestricted protocol view 2»	119
5. 3 Exemples d'images de la même personne de la base de données LFW	118
5. 4 Taux de vérification uni-modale 2D et 3D de visage à 0,001 TFA pour le protocole standard ROC I, II et III (FRGC v2)	137
5. 5 Vérification uni-modale 3D et 2D de visage (CASIA 3D)	138
5. 6 Taux de vérification multimodale du visage à 0,001 TFA pour ROC I, ROC II, ROC III et ROC pour le protocole "neutre contre tout"	147
5. 7 Vérifications multi-descripteurs multimodales de visage (CASIA 3D)	148

θ = Orientation.

λ = Fréquence.

Φ = Phases de la sinusoïde.

σ = Le support temporel.

γ = L'enveloppe de la gaussienne.

N = Résolution de l'image.

m = La moyenne de toutes les images d'apprentissage.

S_w = La matrice de dispersion intra-classes.

m^i = La moyenne de la classe i .

n_i = Le nombre d'échantillons de la classe i .

S_b = La matrice de dispersion inter classes.

\bar{m} = La moyenne de tous les échantillons.

U_{LDA} = La matrice de transformation LDA.

V = La matrice des vecteurs propres.

A = La matrice diagonale des valeurs propres.

K_b = La nouvelle matrice de dispersion inter classe.

U_{EFM} = La matrice de transformation globale de l'EFM.

U_{DLDA} = La matrice de transformation globale de DLDA.

I = Image.

Th = Seuil.

μ = Moyenne arithmétique.

r_1 = Bord gauche de la distribution.

r_2 = Bord droit de la distribution.

μ_A = Vecteur des scores moyens.

Σ = Matrice de covariance entre experts.

$h(x)$ = fonction d'étalement de point.

\angle = Angle de phase.

u = Fréquence.

l = Taille du filtre.

n = Longueur du filtre.

M = Taille du filtre.

W_i = filtres du BSIF.

S_t = Matrice de dispersion totale.

Résumé

La reconnaissance faciale est l'une des meilleures modalités biométriques pour des applications liées à l'identification ou l'authentification de personnes. En effet, c'est la modalité utilisée par les humains. Elle est non intrusive et socialement bien acceptée. Malheureusement, les visages humains sont semblables et offrent par conséquent une faible possibilité de distinction par rapport à d'autres modalités biométriques, comme par exemple, les empreintes digitales et l'iris. Par ailleurs, lorsqu'il s'agit d'images 2D de visages, les variations intra-classe, dues à des facteurs aussi divers que les changements des conditions d'éclairage, variation de cosmétiques et de pose, sont généralement supérieures aux variations inter classes, ce qui rend la reconnaissance faciale 2D peu fiable dans des conditions réelles. Récemment, les représentations 3D de visages ont été largement étudiées par la communauté scientifique pour pallier les problèmes non résolus dans la reconnaissance faciale 2D. Cette thèse est consacrée à la reconnaissance faciale robuste utilisant la fusion des données faciales 2D et 3D.

Nous consacrons la première partie de notre étude à la vérification de visage uni-modale et multi-algorithmes de visage 2D. Tout d'abord, nous étudions plusieurs méthodes pour sélectionner les meilleurs systèmes d'authentification de visages. Ensuite, nous présentons la multi-modalité et les méthodes de fusion de scores pour les deux approches combinaison et classification. Enfin les méthodes de fusion de scores sont comparées sur la base de données des visages XM2VTS.

Dans la deuxième partie nous proposons un algorithme automatique d'authentification du visage par la fusion de deux systèmes multimodaux (multi-algorithmes et multi-capteurs 2D + 3D). Tout d'abord, nous avons corrigé la rotation de la tête par l'algorithme ICP, puis présenté six méthodes d'extraction de caractéristiques locales (MSLBP, CSL proposée, Ondelettes de Gabor, LBP, LPQ et BSIF). La classification des caractéristiques est réalisée par le métrique cosinus après réduction d'espace par EFM, puis fusion au niveau des scores par un bon classificateur à deux classes SVM. Enfin, l'application est réalisée sur les bases de données CASIA 3D et Bosphorus.

Dans la dernière partie, nous étudions la vérification de visage uni-modale 2D et 3D et multimodale (2D+3D) basée sur la fusion de l'information locale. L'étude comprend trois étapes principales (prétraitement, extraction de caractéristiques et classification). Dans la première étape, une phase de prétraitement est nécessaire. L'algorithme ICP est utilisé pour aligner tous les visages et l'approche PS est utilisée pour réduire l'influence de la variation de l'éclairage pour les images 2D. Dans la deuxième étape nous avons utilisé quatre descripteurs locaux (LBP, LPQ, BSIF et *Statistical LBP* proposée). Après extraction des caractéristiques, l'image faciale 2D ou 3D est divisée en 10 régions et chaque région est divisée en 15 petites blocs. L'extraction de caractéristiques locales est résumée par les histogrammes correspondants de chaque bloc. Dans la dernière étape, nous proposons d'utiliser EDA couplée à WCCN pour la réduction de dimension des histogrammes pour chaque région. Nous validons nos méthodes proposées en les comparant avec celles existant dans la littérature scientifique sur les bases de données FRGC v2 et CASIA 3D.

Mots-clés : Reconnaissance faciale 2D et 3D, Multi-modalité, CSL proposée, *Statistical LBP* proposée, EDA, WCCN.

ملخص

التعرف على الوجه هو واحد من أفضل النظم البيومترية المتعلقة بتحديد والتأكد من الأشخاص. في الواقع، هذا هو النظام المستخدم من قبل البشر. فهو غير دخيل، و أيضا اجتماعيا مقبولا. للأسف، الوجوه البشرية متشابهة، وبالتالي دقة التمييز منخفضة بالمقارنة بأساليب البيومترية الأخرى، على سبيل المثال، البصمة الرقمية والقزحية. وعلاوة على ذلك، عندما يتعلق الأمر بصور الوجه ثنائية الأبعاد (2D)، والاختلافات في الصور بالنسبة لنفس الشخص، بسبب اختلافات متنوعة مثل التغيرات في ظروف الإضاءة، والاختلاف في مستحضرات التجميل ودوران الرأس، و هذه العوامل هي أعلى عموما من الاختلافات في الصور بالنسبة لمختلف الأشخاص، الأمر الذي يجعل التعرف على الوجه ثنائي الأبعاد (2D) لا يمكن الاعتماد عليه في ظل ظروف حقيقية. مؤخرا، تمثيلات الوجوه ثلاثية الأبعاد (3D) درست على نطاق واسع من قبل المجتمع العلمي، لمعالجة المشاكل التي لم تحل في التعرف على الوجه (2D). ويخصص هذا البحث على التعرف على الوجه باستخدام مزيج من البيانات الوجه (2D) و (3D).

ننفيق الجزء الأول من دراستنا على التحقق من وجهه (2D) الأحادي و المتعدد الخوارزميات. أولا، ندرس عدة طرق لاختيار أفضل أنظمة التحقق من الوجه. ثم نقدم نظم متعددة الوسائط وطرق دمج النتائج. أخيرا، نقوم بمقارنة طرق إدماج بواسطة قاعدة بيانات الوجوه XM2VTS.

في الجزء الثاني من دراستنا نقترح نظام التحقق من الوجه بواسطة دمج نظامين متعدد الوسائط (متعددة الخوارزميات ومتعددة الاستشعار (2D) + (3D)). أولا، نقوم بتصحيح دوران الرأس باستخدام ICP، ثم نقدم ستة طرق لاستخراج الميزات المحلية (LBP, Statistical LBP, Gabor, LPQ, BSIF). نستعمل EFM لتخفيض من حجم هذه الميزات المحلية، ونستعمل (métrique cosinus) لمقارنة الميزات المحلية المنخفضة الحجم. ثم ندمج النتائج (scores) بالمصنف SVM. وأخيرا، يتم التطبيق باستعمال قاعدات البيانات CASIA 3D و Bosphorus.

في الجزء الأخير، ندرس التحقق من وجه أحادي الوسائط (uni-modale 3D et 2D) و متعدد الوسائط (multimodale (3D+2D))، ويشمل نظامنا ثلاث خطوات رئيسية (المعالجة، استخراج الميزات وتصنيف). في الخطوة الأولى، يتم معالجة صور ثنائية وثلاثية الأبعاد. يتم استخدام ICP لتصحيح دوران الرأس ويستخدم PS للحد من تأثير التغير في الإضاءة للصور ثنائية الأبعاد. في المرحلة الثانية استخدمنا أربعة طرق لاستخراج الميزات المحلية (LBP, LPQ, BSIF et Statistical LBP proposé). يتم تقسيم صورة الوجه ثنائية وثلاثية الأبعاد بعد استخراج ميزت المحلية الى 10 مناطق، يتم تقسيم كل المنطقة إلى 15 كتل (blocs) صغيرة. وتتخلص الميزات المحلية لكل كتلة صغيرة عن طريق (histogramme). في الخطوة الأخيرة نقترح استخدام EDA + WCCN لتخفيض من حجم histogrammes لكم منطقة. وأخيرا، يتم التطبيق باستعمال قاعدات البيانات CASIA 3D و FRGC.

الكلمات الرئيسية: التعرف على الوجه D3 و D2 ، متعددة الوسائط، WCCN, CSL, Statistical LBP, EDA.

Introduction générale

À l'origine, le mot « biométrie » désigne l'application au domaine de la biologie des mesures utilisées en mathématiques. Il renvoie maintenant à un éventail de techniques, d'appareils et de systèmes permettant aux machines de reconnaître des personnes ou de confirmer ou d'authentifier leur identité. Cependant, toutes les caractéristiques d'un être humain ne peuvent pas être utilisées comme moyens d'identification. En effet, pour qu'un système biométrique puisse fonctionner en environnement réel, les caractéristiques physiologiques doivent satisfaire les conditions suivantes [1]:

- **être universelles:** la caractéristique doit être possédée par chaque individu,
- **uniques:** la caractéristique doit permettre la différenciation d'un individu par rapport à un autre,
- **permanentes:** la caractéristique doit être invariante dans le temps,
- **enregistrables:** la caractéristique doit pouvoir être acquise.
- **performante:** la caractéristique doit permettre au système de reconnaître efficacement un individu, en minimisant autant que possible les fausses alarmes,
- **acceptable:** la caractéristique doit être acceptée par les utilisateurs, ce point dépend de la façon dont est perçue la caractéristique par la population,
- **infalsifiable:** la caractéristique doit être difficilement falsifiable afin d'éviter une utilisation frauduleuse du système.

Les différentes caractéristiques par lesquelles il est possible d'identifier un individu sont appelées modalités biométriques physiques ou comportementales. Les plus couramment utilisées (ou étudiées) sont:

- Biométries physiques: l'empreinte digitale, la géométrie de la main, l'empreinte de la paume de la main, l'iris, la rétine, L'ADN (ou Acide Désoxyribo Nucléique), le visage, la géométrie de l'oreille
- Biométries comportementales: la voix, la démarche, la signature, la dynamique de frappe au clavier.

Nombre de ces modalités peuvent de plus donner naissance à d'autres modalités lorsque le capteur est modifié. Par exemple, lors de la capture d'un visage à l'aide d'une caméra infrarouge ou une capture 3D, il s'agit toujours du visage mais on parle alors de modalité infrarouge ou 3D du visage.

Dans cette thèse, nous nous intéressons à la biométrie du visage 2D, puis 3D et enfin multimodale (2D+3D). Le texte est constitué d'une introduction générale, suivie de cinq chapitres et d'une conclusion générale.

Le premier chapitre rappelle des concepts de base et de terminologie en biométrie faciale. Il définit ce qu'est la biométrie en général et justifie le choix de la reconnaissance du visage en particulier. Il évoque les difficultés de reconnaissance de visage 2D et énumère les bases de données 2D de visage disponibles. Il introduit ensuite la reconnaissance de visage 3D à travers la présentation des techniques 3D et la donnée des bases de données 3D de visage disponibles.

Le deuxième chapitre résume l'état de l'art sur les deux aspects complémentaires de nos travaux: d'une part la reconnaissance de visages 2D et, d'autre part, la reconnaissance faciale 3D. Nous commençons ce chapitre par un examen des importants algorithmes de reconnaissance de visage 2D, car la plupart des algorithmes de reconnaissance de visage 3D et 2D+3D sont des extensions directes d'algorithmes 2D existants. Le but de ce chapitre est d'évaluer les capacités et les lacunes d'algorithmes de reconnaissance de visage en cours afin d'identifier les problèmes et les orientations pour la recherche future dans ce domaine.

Le troisième chapitre est consacré à notre approche de vérification uni-modale et multi-algorithmes de visage 2D. Dans la partie de vérification uni-modale de visage 2D, nous présentons tout d'abord notre modèle de vérification de visage. Puis, nous détaillons les quatre principales étapes de ce modèle. Enfin nous commençons l'étude expérimentale sur la base de données XM2VTS par la comparaison des algorithmes de réduction d'espace de données, puis nous étudions l'extraction des caractéristiques par les ondelettes de Gabor. Dans la partie de vérification de visage 2D multi-algorithmes, nous passons en revue les limitations des systèmes biométriques lorsqu'ils n'utilisent qu'une seule modalité biométrique. Ensuite, nous présentons les différentes multi-modalités possibles, et les méthodes de fusion de scores. Enfin nous détaillons les résultats expérimentaux

Dans le quatrième chapitre, nous examinons comment les caractéristiques locales de l'information 2D et 3D contribuent à la reconnaissance faciale lorsque l'éclairage, les changements d'expressions et la combinaison d'expression sous l'éclairage sont pris en compte. Nous présentons, dans un premier temps, la phase de prétraitement: par l'alignement (translation et rotation) de l'image d'entrée 3D par rapport à l'image de référence (image de visage d'expression neutre) en utilisant l'algorithme ICP (Iterative Closest Point). Selon les images alignées, nous obtenons les images de

profondeur 3D et des images d'intensité 2D. Ensuite, nous détaillons les méthodes d'extraction de caractéristiques utilisées et notre descripteur proposé CSL (Caractéristiques Statistiques Locales), puis la méthode de réduction d'espace basée sur l'EFM et notre système de fusion de scores. Enfin, nous testons notre approche sur les bases de données: CASIA 3D et Bosphorus.

Le cinquième chapitre présente un système de vérification de visage multimodal efficace basé sur la fusion de l'information locale. Dans la première étape, une phase de prétraitement est nécessaire pour remplissage des trous et le ré échantillonnage des données 3D et 2D. L'algorithme ICP (Iterative Closest Point) est utilisé pour aligner tous les visages 2D et 3D avec le premier visage 3D (expression neutre) et l'approche PSA (preprocessing sequence approach) est utilisée pour réduire l'influence de la variation de l'éclairage pour les images 2D. Pour cela, nous utilisons quatre descripteurs locaux. L'image faciale 3D et 2D est divisée en 10 régions et chaque région est divisée en 15 petits blocs. Pour chacune des méthodes utilisées et pour chaque bloc, l'extraction de caractéristiques locales est résumée par les histogrammes correspondants. Les histogrammes des différents blocs appartenant à la même région sont concaténés pour former un vecteur de caractéristiques représentant un descripteur spécifique à une échelle spécifique. Pour la réduction de dimensionnalité, nous proposons d'utiliser EDA (Exponential Discriminant Analysis) couplé à WCCN (Within Class Covariance Normalization) pour réduire l'effet des directions de haute intervariabilité et renforcer la discrimination pour chacune des régions. Enfin, les scores des différentes régions correspondant aux différents descripteurs à différentes échelles sont normalisés par la normalisation de scores Z_{norm} et sont fusionnés par simple somme. Enfin nous validons nos méthodes proposées et nous comparons ces méthodes avec les méthodes existantes dans l'état de l'art sur les bases de données FRGC v2 et CASIA.

Nous terminerons la rédaction de cette thèse sur une conclusion générale sur ces travaux et présenterons les perspectives futures envisageables.

1.1 Introduction

Nous allons commencer par rappeler des concepts de base et de terminologie en biométrie faciale en définissant ce qu'est la biométrie en général et en justifiant le choix de la reconnaissance du visage en particulier. Nous évoquerons ensuite les difficultés de reconnaissance de visage 2D et énumérerons les bases de données 2D de visage disponibles. Nous introduirons enfin la reconnaissance de visage 3D à travers la présentation des techniques 3D et la donnée des bases de données 3D de visage disponibles.

1.1.1 Qu'est-ce que la biométrie?

Un système biométrique peut être soit un système d'identification (reconnaissance) ou un système de vérification (authentification).

L'identification (reconnaissance) est un type d'application pour lequel le système doit répondre à la question : Qui suis je ?. Le système doit trouver l'identité d'une personne parmi celles d'une base de données contenant des personnes déjà enrôlées et renvoyer l'identité correspondant à la personne se présentant devant le système, ou l'identité « inconnue » si cette personne ne fait pas partie de la base. Il s'agit ici d'une comparaison 1 à n où n représente le nombre de personnes dans la base, également appelée galerie. Parmi les utilisations possibles d'un système en mode d'identification, on retrouve la recherche d'individus dangereux, ou encore l'accès restreint d'un bâtiment d'une entreprise à ses seuls employés par exemple.

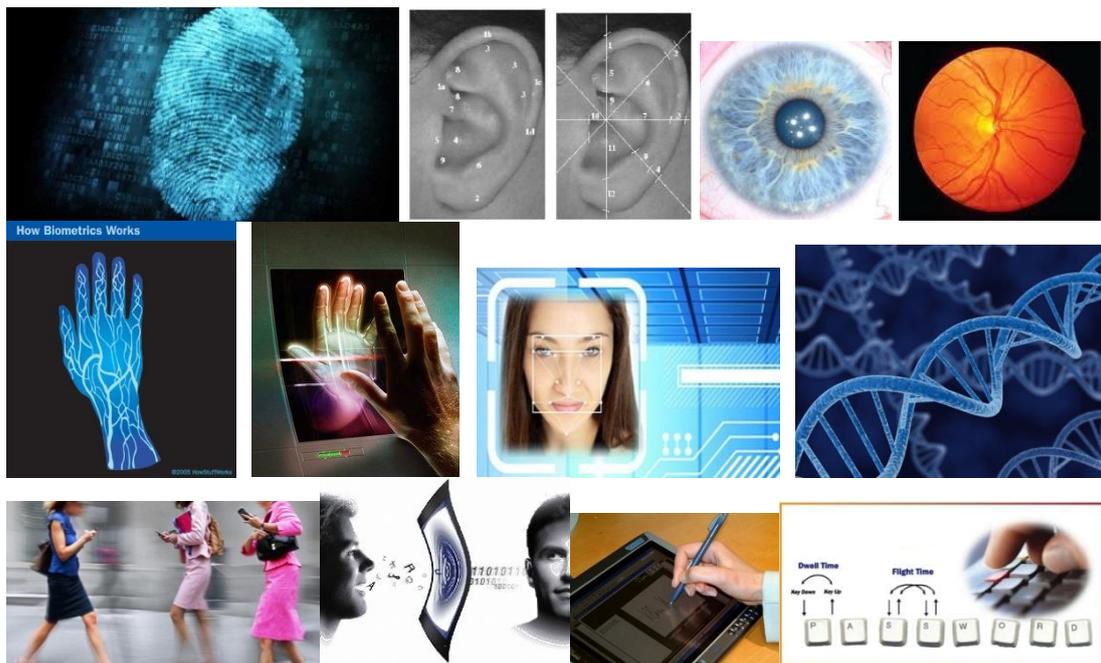


Figure 1.1: Quelques modalités biométriques (L'image est recueillie à partir d'Internet).

L'authentification (vérification) est un type d'application pour lequel le système doit répondre à la question : Suis je bien la personne que je prétends être ?. Le cas d'usage est une personne clamant son identité au système et celui-ci doit alors vérifier si la personne est bien la personne qu'elle prétend être. Parmi les applications liées à l'authentification, citons l'accès à des données sécurisées, des ressources informatiques ou encore des transactions sécurisées.

Les paramètres suivants sont utilisés pour la mesure de la performance standard d'un système biométrique d'identification.

- 1) **Taux de reconnaissance (Rank-one Recognition Rate):** Il mesure le pourcentage des entrées qui sont correctement identifiées.
- 2) **Cumulative Match Characteristic (CMC):** La courbe CMC donne le pourcentage de personnes reconnues en fonction d'une variable que l'on appelle le rang [2]. On dit qu'un système reconnaît au rang 1 lorsqu'il choisit la plus proche image comme résultat de la reconnaissance. On dit qu'un système reconnaît au rang 2, lorsqu'il choisit, parmi deux images, celle qui correspond le mieux à l'image d'entrée, etc. On peut donc dire que plus le rang augmente, plus le taux de reconnaissance correspondant est lié à un niveau de sécurité faible.

Les paramètres suivants sont utilisés pour la mesure de la performance standard d'un système biométrique avec un scénario de vérification.

- 1) **Taux de Faux Rejet (TFR) ou (False Reject Rate, FRR):** ce taux représente le pourcentage de personnes censées être reconnues mais qui sont rejetées par le système.
- 2) **Taux de Fausse Acceptation (TFA) ou (False Accept Rate, FAR):** ce taux représente le pourcentage de personnes censées ne pas être reconnues mais qui sont tout de même acceptées par le système.
- 3) **Taux d'Égale Erreur TEE ou (Equal Error Rate, EER):** Ce taux est calculé à partir des deux premiers critères et constitue un point de mesure de performance courant. Ce point correspond à l'endroit où $TFR = TFA$, c'est-à-dire le meilleur compromis entre les faux rejets et les fausses acceptations.
- 4) **Receiver Operating Characteristic (ROC):** La courbe ROC est une démonstration visuelle du compromis entre les TFA et le TFR liées à un seuil variable.

Comme il est souligné dans le livre "Handbook of Biometrics" [3], la biométrie est un domaine qui évolue rapidement avec largement des applications allant de l'accès en toute sécurité à son ordinateur pour gagner l'entrée dans un pays. Pour plus d'informations, on peut se référer aux manuels [3, 4].

1.1.2 Pourquoi la reconnaissance de visage?

Compte tenu de la nécessité de déterminer l'identité des personnes, la question évidente est qu'est ce qui est le mieux placé pour fournir cette information [5]? Parmi tous les traits biométriques disponibles, les empreintes digitales, l'iris, le visage, la voix et la signature ont gagné large attention. En particulier, les empreintes digitales, l'iris et des techniques de reconnaissance faciale sont à ce jour devenus progressivement les technologies de reconnaissance biométrique grand public. Bien que les techniques de reconnaissance d'empreintes digitales et de l'iris puissent atteindre des résultats fiables dans certaines applications commerciales particulières, elles ont les inconvénients évidents suivants [6]:

- 1) **Physiquement intrusives:** l'empreinte digitale nécessite la coopération de l'utilisateur à entrer en contact physique avec la surface du capteur, tandis que l'iris exige de l'utilisateur de positionner l'œil par rapport au capteur. En outre, ce genre de techniques d'analyse de coopération requiert également de l'utilisateur de faire une pose pendant une seconde pour «déclarer» eux-mêmes [5].
- 2) **Socialement intrusives:** les gens ne peuvent pas reconnaître d'autres personnes utilisant ce type de données, ces types d'identification n'ont pas de place dans les interactions humaines normales et des structures sociales [5].

Pour construire un magasin qui reconnaît ses meilleurs clients, ou un kiosque d'information qui se souvient de vous, ou une maison qui connaît les gens qui y vivent, la reconnaissance des visages de la vidéo et la reconnaissance vocale ont une place naturelle dans ces environnements intelligents de prochaine génération [5]. En particulier, elles doivent être:

- 1) **Naturelles et non-intrusives:** elles sont discrètes (capables de reconnaître à distance) et généralement passives (ne nécessitent pas de générer un éclairage électro magnétique spécial). Elles ne doivent pas restreindre le mouvement de l'utilisateur et doivent être de faibles puissance et peu coûteuses [5].
- 2) **Perception biologique:** C'est peut-être le plus important. Cependant, les humains qui identifient d'autres personnes par leur visage et leur voix sont donc susceptibles d'être à l'aise avec les systèmes qui utilisent le visage et la reconnaissance vocale [5].

Comme une preuve solide de l'organisation ICAO (International Civil Aviation Organization), Hietmeyer [7] a souligné que l'identification biométrique peut permettre un traitement rapide et sécurisé des passagers aériens. Pour sélectionner une seule caractéristique biométrique pour une utilisation dans la confirmation d'identité assistée par ordinateur, il a suggéré l'évaluation de la compatibilité de six traits biométriques: le visage, empreintes digitales, géométrie de la main, la voix, les yeux ainsi que la signature sur la base d'un système MRTD (Machine Readable Travel Documents). Le score de compatibilité implique: les facteurs de l'inscription, le renouvellement des données, les exigences de vérification de l'identité assistée par machine, la redondance, la perception du public, les besoins de stockage et de la performance. Comme le montre la figure 1.2, le système de reconnaissance de visage a marqué la plus grande compatibilité et est en train de devenir la biométrie la plus susceptible d'être sélectionnée pour une utilisation internationale.

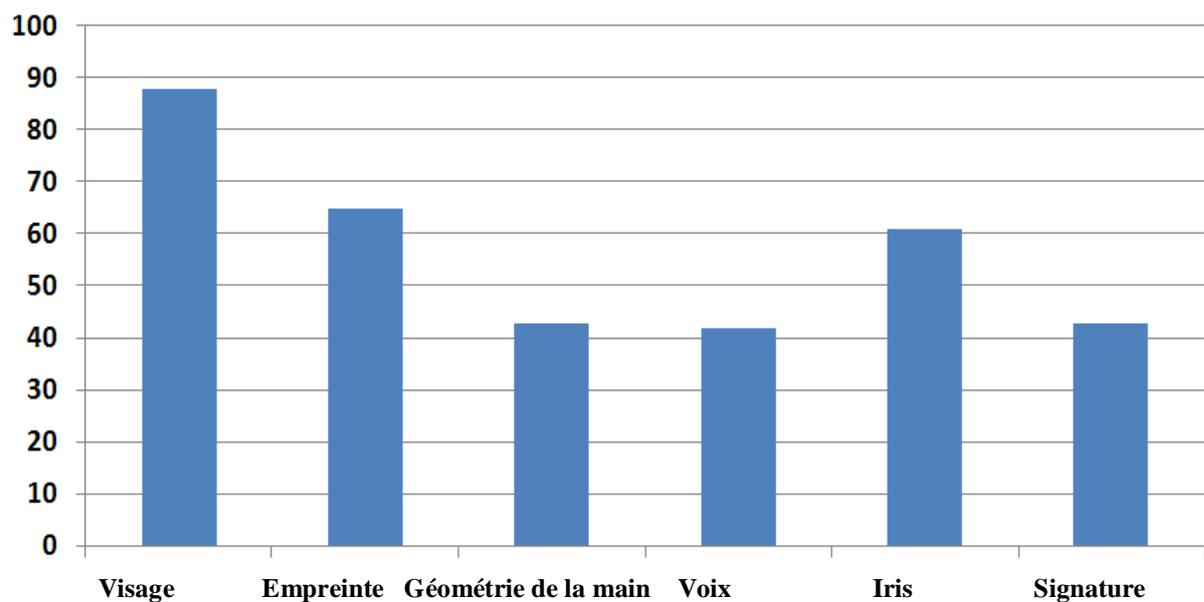


Figure 1.2: Comparaison de compatibilité MRTD pour diverses caractéristiques biométriques [7].

1.1.3 Difficultés de reconnaissance de visage 2D

De nombreuses propriétés du visage ainsi que les conditions dans lesquelles elles ont été photographiées rendent le traitement automatique difficile. Dans le cadre de la reconnaissance, le principal problème sous-jacent est la variance intra-classe, c'est à dire la variabilité que peut prendre le visage d'une même personne à cause de différences de luminosité, de pose . . . Cette variation intra-classe peut être supérieure à la variance inter classe, c'est à dire la variabilité que prennent les visages de différentes personnes. Dans de nombreux systèmes, cette variation intra-classe est considérée comme du bruit (information non désirée) rendant l'objectif de la reconnaissance plus difficile. L'extraction de caractéristiques discriminantes est en effet rendue plus compliquée et les performances globales des systèmes s'en trouvent amoindries [8]. Nous détaillons ici les principales difficultés rencontrées par un système de reconnaissance faciale automatique 2D dans des conditions réelles.

1.1.3.1 Illumination

Les variations d'illumination viennent entraîner des variations considérables dans l'apparence d'un visage. Deux types d'éclairage peuvent influencer celle-ci: l'illumination globale (ou ambiante) et l'illumination locale. Alors que l'illumination globale affecte tout le visage de manière uniforme (ou presque), l'illumination locale entraîne la création de zones d'ombre et de zones éclairées et ce de manière non linéaire. La figure 1.3 présente un exemple de visage dont la source lumineuse l'éclairant se déplace. De nombreuses approches ont été proposées pour gérer ces problèmes de luminosité. Une modélisation implicite de la luminosité dans la création d'un modèle de visage peut être réalisée. L'extraction de caractéristiques invariantes aux changements de luminosité est également une approche largement décrite dans la littérature. Notons enfin qu'un certain nombre d'approches traitent le problème de la luminosité en amont de la reconnaissance par une étape de prétraitement dont l'objectif principal est bien souvent de corriger les artefacts dû aux variations de luminosité.



Figure 1.3: Exemple d'un visage d'une même personne subissant un changement de luminosité (L'image est recueillie à partir d'Internet).

1.1.3.2 Pose

La pose d'un visage définit la rotation qu'a pu subir un visage lors de la capture. Les variations de pose peuvent être de deux types selon le type de rotation: rotation dans le plan où l'axe de rotation est l'axe de la caméra et rotation hors-plan sinon. La figure 1.4 présente un exemple d'un visage subissant une rotation hors plan. Les variations de pose affectent grandement les systèmes de reconnaissance automatique de visages, c'est pourquoi nombre d'entre eux se limite au pose frontales, ou à des pose spécifiques requérant cependant une estimation au préalable. Dans le cas d'une rotation dans le plan, l'apparence du visage n'est pas déformée et une bonne estimation de l'angle de rotation peut suffire à recalculer l'image par simple rotation inverse et ainsi obtenir une pose frontale (front en haut de l'image, menton en bas). Le cas de la rotation hors plan est souvent bien plus complexe, sauf si les visages utilisés pour l'enrôlement et la reconnaissance présentent la même pose.

1.1.3.3 Expressions faciales

L'apparence d'un visage varie grandement en présence d'expressions faciales (figure 1.5). Les éléments faciaux tels que la bouche ou encore les yeux peuvent alors subir des déformations importantes, pouvant faire échouer un système de reconnaissance faciale fondée par exemple sur des points d'intérêt (ceux-ci pouvant ainsi subir d'importantes translations). La bouche est en général l'élément facial qui varie le plus, mais l'aspect des sourcils peut par exemple être grandement modifié [9].



Figure.1.4: Exemple d'un visage d'une même personne subissant des variations de pose hors plan (Image recueillie à partir d'Internet).

1.1.3.4 Oclusions

Les oclusions partielles apparaissent fréquemment dans des applications réelles, comme illustré sur la figure 1.6. Elles peuvent être causées par une main cachant une partie du visage, par des cheveux longs, des lunettes de vue, de soleil, par tout autre objet (foulard . . .), ou encore par une autre personne. Il arrive également qu'une partie du visage en cache une autre, comme dans le cas d'une rotation hors plan par exemple.

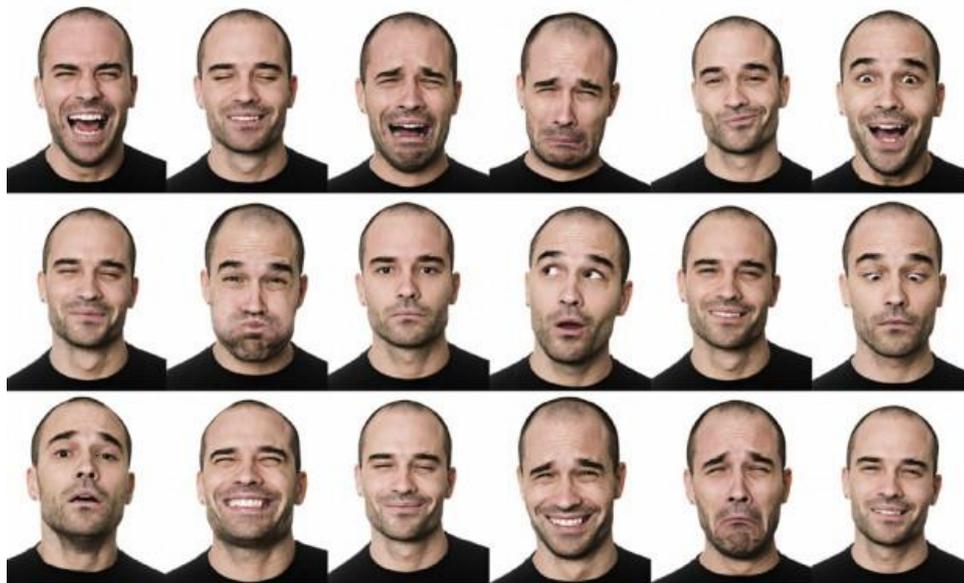


Figure.1.5: Variabilité intra-classe due à la présence d'expressions faciales (Image recueillie à partir d'Internet).



Figure.1.6:Exemples d'occlusion du visage (Image recueillie à partir d'Internet).

1.1.3.5 Autres difficultés

D'autres types de difficultés peuvent apparaître pour un système automatique de reconnaissance faciale. Ainsi, des variations peuvent être dues à la présence de maquillage, d'opérations chirurgicales, de différentes coupes de cheveux, ou encore la présence (absence) de moustaches, de barbes, . . . Un autre point très important est l'âge des captures, c'est à dire le moment auquel les captures ont été réalisées. En effet, l'apparence d'un visage peut changer au cours du temps (notamment lors de l'adolescence) et un écart de temps important entre deux captures peut engendrer des difficultés de reconnaissance. Cet effet n'est pas nouveau mais la constitution de bases de données s'étalant sur plusieurs années est difficile.

1.1.4 Bases de données 2D de visage

De nombreuses bases de données de visages (publiques ou privées) existent à des fins de recherche. Elles peuvent différer entre elles sur plusieurs points :

- le nombre d'images disponibles est probablement le critère le plus important d'une base de données,
- le nombre d'images disponibles par personne,
- la taille des images,
- la pose et orientations des visages,
- la variation de l'illumination,
- le sexe des personnes présentes,
- la présence d'artefacts (lunettes, barbes, . . .),
- la présence d'images statiques ou de vidéos,
- la présence d'un fond uniforme,
- la période entre les prises de vues.

Il est ainsi recommandé de bien choisir la base de données lors des tests d'un algorithme. En effet, certaines sont dotées d'un protocole bien défini permettant ainsi la comparaison directe des résultats. De plus, le choix doit dépendre du problème que l'on souhaite tester: illumination, reconnaissance à travers le temps, expressions faciales.... La disponibilité de nombreuses images différentes par personne peut également être un argument décisif pour la bonne réalisation d'un algorithme.

Le tableau 1.1 présente les principales bases de données de visages 2D. Ces bases de données varient en: image RVB ou gris, taille, nombre de personnes, nombre de photos par personnes, variations de l'image (illumination (i), pose (p), expression (e), occlusions (o), retard de temps (t)) et page d'accueil sur le Web.

Afin d'assurer une reproductibilité des expériences et la comparaison avec d'autres méthodes, nous testons notre approche dans la partie de reconnaissance de visage 2D de notre travail sur les bien-connues bases de données 2D XM2VTS et LFW en utilisant les protocoles communs. La base de données XM2VTS est une base contrôlée et la base de données LFW est une base non contrôlée qui contient des variations: d'éclairage, de pose, d'expression, d'occlusion et de retard de temps.

Tableau 1.1: Base de données 2D. Variations de l'image sont indiqués par (i) l'éclairage, (p) pose, (e) l'expression, (o) occlusion, (t) retard de temps.

Nom	RVB /gris	Taille de l'image	Nombre de personnes	Photos / personne	Variations de l'image	Adresse web
AR visage	RVB	576×768	126 70Mâles 56 Femelles	26	i, e, o, t	[10]
CVL	RVB	640×480	114 108 Mâles 6 Femelles	7	p, e	[11]
Yale visage B	gris	640×480	10	576	p, i	[12]
Yale visage	gris	320×243	15 14 Mâles 1 Femelles	11	i, e	[13]
PIE	RVB	640×486	68	~608	p, i, e	[14]
UMIST visage	gris	220×220	20	19–36	p	[15]
Olivetti Att – ORL	gris	92×112	40	10		[16]
JAFFE	gris	256×256	10	7	e	[17]
The Human Scan Database	gris	384×286	23	~66		[18]
The University of Oulu Physics-Based Face Database	gris	428×569	125	16	i	[20]
XM2VTS	RVB	576×720	295		p	[20]
FERET	RVB	256×384	30000		p, i, e, t	[21]
LFW	RVB	150×150	13233		p, i, e, o, t	[22]

1.1.5 Opportunités de reconnaissance de visage 3D

Il y a tant de difficultés techniques pour la reconnaissance faciale 2D du visage. On peut suggérer d'utiliser les données 3D pour la reconnaissance faciale. Cette suggestion déclenche naturellement une question: «Pourquoi la reconnaissance faciale 3D ?». Pour répondre à cette question, nous listons plusieurs raisons principales.

- 1) L'invention et la vulgarisation des techniques de numérisation 3D: avec l'invention et le développement rapide des technologies de numérisation 3D, le monde connaît une révolution numérique 3D. Le mesurage, la reconstruction, la numérisation et la reproduction d'objets en 3D du monde réel sont de plus en plus faciles et populaires. De nos jours, de nombreux types de capteurs 3D sont disponibles pour capturer des données 3D statiques et même dynamiques de surface du visage. On peut se référer à la section 1.2.1 «Techniques d'acquisition de visage 3D » pour des exemples détaillés.
- 2) La disponibilité d'un grand potentiel pour lutter contre les difficultés 2D: les informations sur la forme des approches fondées sur la reconnaissance faciale en 3D sont naturellement moins sensibles aux conditions d'éclairage par comparaison avec les informations de photométrie 2D. La surface faciale 3D est plus commode pour effectuer la correction de pose et résoudre le

problème de rotation du hors-plan dans la reconnaissance faciale 2D. En outre, les cosmétiques pour le visage ont tendance à avoir moins d'impact pour les scans du visage 3D [23].

- 3) La forme du visage 3D et la photométrie 2D contiennent des informations comparables et complémentaires: Comme indiqué dans [24], de nombreux résultats de l'étude fournissent la preuve que les indices photométriques 2D et de forme 3D sont utilisés à peu près par l'homme pour reconnaître les visages. En outre, comme indiqué dans [25], nous sommes particulièrement intéressés à la reconnaissance faciale 3D car il est communément admis que l'utilisation de la détection 3D a le potentiel pour une plus grande précision de la reconnaissance qu'en 2D. D'autre part, il a été prouvé à maintes reprises que les performances de la reconnaissance de visage multimodale (2D +3D) sont toujours meilleures que celles en 2D seul [25, 26]. C'est-à-dire, la forme du visage et la photométrie sont aussi complémentaires les unes aux autres pour la reconnaissance faciale.

1.2 Reconnaissance de visage 3D

1.2.1 Techniques d'acquisition de visage 3D

Dans la dernière décennie, les techniques de numérisation, de capture et de reconstruction de visages 3D sont témoins d'un formidable développement. Dans cette section, nous résumons trois types de techniques [6]: lumière structurée « structured light », multi-vues stéréo « multi-view stereo » et stéréo photométrie « photometric stereo ».

- 1) **Lumière structurée (structured light):** En tant que technique de balayage "active", la lumière structurée a été largement utilisée dans l'acquisition de données de visages 3D [27, 28, 29, 30, 31]. La configuration de base de cette technique comprend une caméra et un projecteur de lumière dans les relations géométriques données. Le projecteur projette un ou plusieurs motifs de lumière codés sur la scène. La forme des signaux sont extraits par la mesure de la déformation de motifs lumineux à la surface de l'objet. La figure 1.7 illustre un exemple des motifs de lumière codés et projetés sur un visage. Actuellement, un scanner 3D structuré à base de lumière, peut capturer des séquences statiques ou des scans du visage en 3D ainsi que des images 2D du visage en temps réel, ou même à haute vitesse. L'inconvénient est que les visages des utilisateurs devraient être limités dans une quantité limitée de la zone couverte non seulement par la lumière projetée structurée mais aussi visible sur la caméra. Certains scanners 3D populaires basés sur des techniques d'éclairage structurés sont « Minolta Vivid 900/910 series » [32], « Inspeck Mega Capturor II 3D » [33], et « Kinect » [34], etc.



Figure 1.7: Exemple de lumière structurée (à gauche) et sa projection sur un visage [35].

- 2) **Multi-vues stéréo (Multi-view stereo):** Comme une technique « passive » la capture multi-vues stéréo a également été largement utilisée dans l'acquisition de données faciales 3D [36]. Cette

technique utilise plusieurs caméras calibrées placées à différents points de vue du sujet. Sous diverses contraintes, les points correspondants de différentes images de bon sens peuvent être mis en place et ainsi la forme du visage peut être reconstruite. Cette technique ne nécessite pas de clignotant, puisque tout l'appareil peut enregistrer le même sens en même temps avec des lumières constantes. Cependant, la reconstruction précise des surfaces lisses (par exemple certaines régions du visage sans beaucoup de caractères de texture) est très difficile en utilisant cette technique [25]. En outre, la reconstruction 3D doit être utilisée hors ligne en raison de l'énorme complexité des calculs impliquant cette technique. Des systèmes commerciaux populaires qui emploient la technique de stéréo multi-vues sont le système de capture de visage dynamique DI3D [37] et le système stéréo 3D dynamique 3DMD [38]. Les deux systèmes ont été utilisés pour l'acquisition de bases de données de visage 3D [39, 40, 41]. *Beeler et al.* [42] ont développé un système prometteur de capture de visage 3D basé sur la technique de stéréo multi-vues. Ce système permet de reconstituer la surface du visage avec une précision submillimétrique, ce qui est comparable avec ces systèmes à la pointe de la technologie basée sur la lumière structurée [42]. La configuration de base et des visages reconstruits par ce système sont représentés dans la figure 1.8.

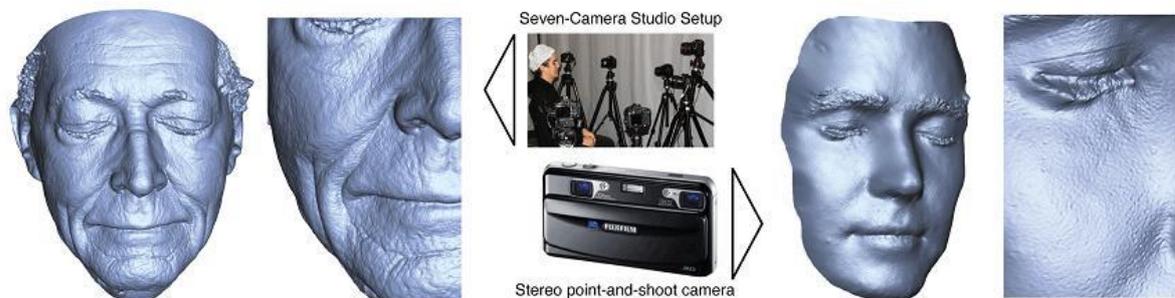


Figure 1.8: Gauche: modèle de visage capturé en utilisant une configuration de studio de sept caméras, centre: systèmes de capture, droite: modèle de visage capturé avec la caméra stéréo binoculaire des consommateurs [42].

- 3) **Photométrie Stéréo (Stereo photometric):** La photométrie stéréo, introduite à l'origine par *Woodham* en 1980 [43], est une technique pour estimer les normales de la surface d'objets en capturant un ensemble d'images de l'objet sous différentes conditions d'éclairage. La figure 1.9 montre quelques reconstructions produites par quatre feux standards photométrie stéréo et la stéréo photométrique multi-spectrale [44]. Par exemple, le système de capture de données 3D stéréo photométrique développé par *Hansen et al.* [45] est capable d'acquérir quatre images brutes en environ 20 ms. Cependant, cette technique est très sensible à la présence d'ombres, aux faits saillants et à l'éclairage non uniforme. En outre, au lieu d'obtenir des nuages de points ou des surfaces maillées directement, la technique photométrie stéréo estime que les normales de la surface de l'objet et la procédure d'intégration, suite à la reconstruction de surface, ajoute plus de temps de calcul et introduit également des erreurs supplémentaires [46]. Un système stéréo photométrique très récemment développé et utilisé pour l'acquisition de données de visages 3D est introduit par [47].

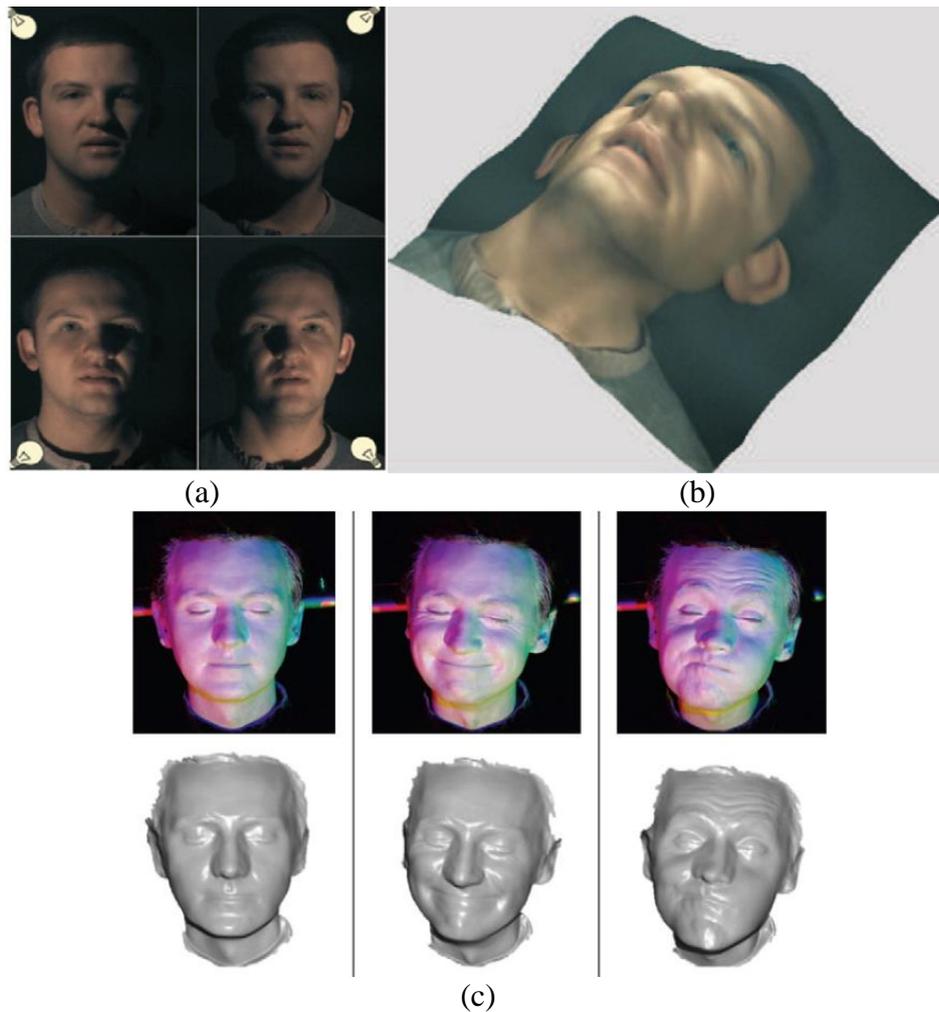


Figure 1.9: Comparaison de la stéréo photométrie et stéréo photométrie multi spectrale. **(a)** Procédé d'éclairage photométrique stéréo. **(b)** Données faciales 3D capturées par photométrie stéréo. **(c)** Stéréo photométrie multi spectrale (capture de l'objet avec une lumière de couleur différente dans une direction différente). Photo est prise à partir de [44].

1.2.2 Concepts de base et terminologie

Depuis que l'analyse de visage 3D est une caractéristique biométrique, tous les concepts et la terminologie du système général de reconnaissance biométrique sont adaptés à la reconnaissance de visage 3D à l'exception de quelques cas. Pour être plus précis, nous donnons un bref résumé dans cette section. La reconnaissance générale de visage 3D fait référence à deux catégories d'applications. Une application est appelée reconnaissance faciale 3D ou identification de visage 3D. L'autre est appelée vérification de visage 3D ou authentification. L'identification de visage 3D est un problème de correspondance un-à-plusieurs alors que la vérification de visage 3D est un problème de correspondance un-à-un. Le taux de reconnaissance et la courbe CMC (Cumulative Match Characteristic) sont couramment utilisés pour évaluer l'identification de visage 3D. Le TEE (Taux d'Égale Erreur) et la courbe ROC sont couramment utilisés pour évaluer la vérification de visage 3D. Dans les deux cas, un ensemble de scans du visage 3D de sujets connus sont initialement existes. Cet ensemble est généralement appelé ensemble de galerie (gallery). Lors de la reconnaissance, les scans du visage 3D des sujets appartiennent à l'ensemble de la galerie où d'autres sujets sont appelés sondes (probes), qui sont utilisés pour correspondre avec les scans dans l'ensemble de galerie. Dans notre travail nous étudions la vérification de visage 3D car c'est l'application la plus difficile.

La forme et la textures sont deux propriétés intrinsèques d'un objet du visage. Heureusement, le système actuel d'acquisition de visage 3D peut généralement offrir des scans du visage 3D (forme) avec leurs textures correspondantes. Dans ce cas, une image 2D est aussi appelée « image 2D de texture ». Un visage de numérisation 3D avec sa carte de texture est généralement appelé « visage 3D texturé ». Une analyse de visage 3D peut être rendue sous forme d'image de profondeur, de nuages de points ou de maillage triangulaire. L'image de profondeur est une image dans laquelle la valeur du pixel enregistre la distance entre le capteur et la surface du visage. Les nuages de points enregistrent la forme de la géométrie d'un modèle de visage 3D en coordonnées 3D. Le maillage triangulaire enregistre la forme de la géométrie d'un modèle de visage 3D en points 3D et leurs structures de topologie. La figure 1.10 illustre un exemple de l'ensemble de ces quatre types de formats d'enregistrement de visage.

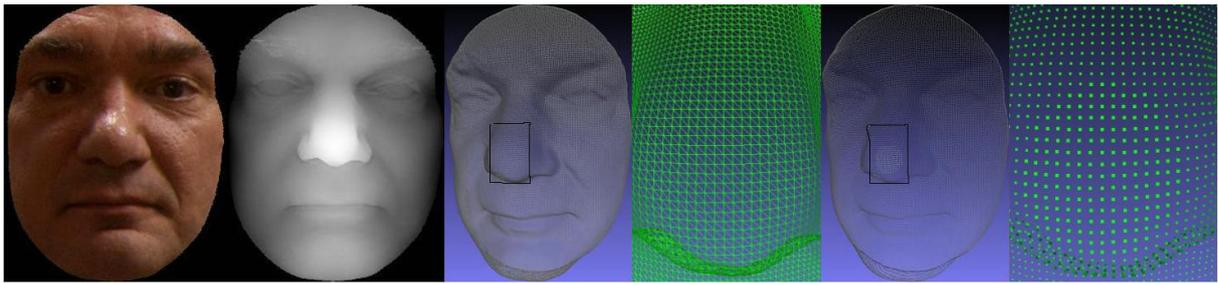


Figure.1.10: Quatre types de formats d'enregistrement du visage 3D, de gauche à droite: texture de visage 3D, image de profondeur, maille en forme de triangle et la zone de nez zoomé, nuages de points et la région de nez zoomée [6].

Selon les modalités de visage adopté pour la galerie et la sonde, la reconnaissance faciale 3D peut être effectuée dans trois scénarios correspondants: reconnaissance de visage uniquement 3D de forme à base (forme 3D contre forme 3D), reconnaissance faciale multimodale 2D+3D (3D texturé contre 3D texturé) et la reconnaissance de visage 3D asymétrique (forme 3D contre texture 2D ou vice versa). La reconnaissance de visage multimodale (2D+3D) s'est avérée être plus précise que la reconnaissance uniquement 3D de forme à base de visage [25]. Dans ce travail, nous nous concentrons sur le scénario de reconnaissance faciale multimodale 2D+3D.

1.2.3 Base de données 3D de visage

De nombreux organismes de recherche ont construit différentes bases de données de visage 3D pour évaluer les algorithmes de reconnaissance faciale 3D, par exemple: XM2VTS [48], 3D_RMA [49], MPI [50], York [51], Notre Dame [52], Gavab [52], BU-3DFE [53], BU-4DFE [39] ...etc. Pour une introduction aux bases de données de visage 3D, on peut consulter le tableau 1.2. La figure 1.11 présente un balayage de chacune des bases de données du visage les plus utilisées en 3D contenant des variations d'expression.

D'après le tableau 1.2, nous pouvons voir que chaque base de données faciale 3D a été conçue pour un ou plusieurs problèmes spécifiques de recherche dans ce domaine:

- Modèles d'un grand nombre et variété démographique de sujets: FRGC v2.
- Modèles d'une personne donnée à intervalles répétés de temps: FRGC v2.
- Modèles avec des variations importantes d'expressions faciales: BU-3DFE, Bosphorus, FRGCv2.0, CASIA, York, etc.
- Les modèles de jumeaux: 3DTEC.

Dans notre travail, trois bases de données sont utilisées pour les évaluations et les analyses expérimentales sur notre approche proposée, à savoir CASIA, Bosphorus et FRGC v2. Les chapitres 4 et 5 expliciteront cette approche.

Tableau 1.2: Quelques bases de données 3D de visage. Variations du visage indiquées par (i) illumination, (p) pose , (e) expression, (o) occlusion, (t) retard de temps.

Nom	Appareil	Nombre de personnes	Photos / personne	Texture	Variations de visage
XM2VTS [48]	-	295	1	Oui	-
3D_RMA [49]	-	120	3	Non	e, p, o
MPI [50]	Cyberware	200	7	-	-
York [51]	-	350	15	Oui	e, p, t
Notre Dame [52]	Minolta Vivid 910	277	Total:953	Oui	e, t
Gavab [52]	Minolta Vivid 700	61	9	Non	p, e
BU-3DFE [53]	3DMD	100	4	Non	e
BU-4DFE [39]	-	101	Total: 60600	Oui	e
Bosphorus [54]	InspeckMegaCapturor II	105	Total:4666	Oui	e, p, o
FRGC v1 [55]	Minolta Vivid 910	275	Total:943	Oui	-
FRGC v2 [55]	Minolta Vivid 910	466	Total:4007	Oui	e, o, i, t
FRAV3D [56]	Minolta Vivid400	106	16	Oui	e, p
3DTEC [57]	Minolta Vivid 910	214	2	Oui	e, jumeau
USF 3D [57]	Cyberware 3030 PS	100	1	Oui	-
BJUT-3D [58]	Cyberware 3030 PS	100	-	Oui	e
Biometrics [59, 60]	Minolta Vivid 900	275	Total:1906	Oui	-
CASIA [61]	Minolta Vivid 910	123	Total:4624	Oui	e, p, o, i
Texas 3DFRD [62]	Atlanta, GA	118	Total:1149	Oui	-
MSU [63]	Minolta Vivid 910	90	Total: 533	Non	e, p
ZJU-3DFED [64]	InSpeck 3D MEGA Capturor DF	40	9	Oui	e
SHREC'08 [65]	Minolta Vivid 700	61	Total: 427	Non	e, p

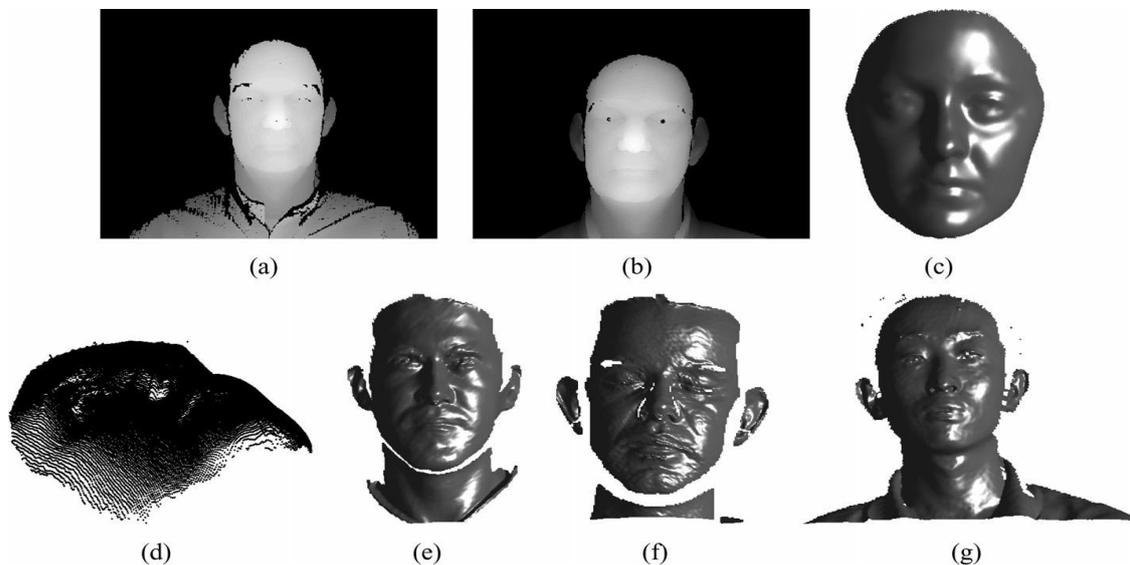


Figure 1.11: Balayage de chacune des bases de données du visage les plus utilisées en 3D contenant des variations d'expression: (a) FRGC v1, (b) FRGC v2, (c) BU-3DFE, (d) Bosphorus, (e) SHREC'08, (f) FRAV and (g) CASIA [66].

1.3 Conclusion

Dans ce chapitre, nous avons rappelé les définitions des deux applications de la biométrie : l'identification et la vérification. Nous avons rappelé également les paramètres utilisés pour l'évaluation des performances de ces deux applications. Nous avons détaillé ce qu'est la biométrie en général et justifié le choix de la reconnaissance du visage en particulier. Nous avons évoqué les limites de la reconnaissance de visage 2D et présenté ensuite la reconnaissance de visage 3D à travers la présentation des techniques 3D.

2.1 Introduction

La reconnaissance par la machine de visages humains a été le but de nombreux chercheurs depuis le projet "homme-machine" de *Bledsoe* en 1960 [67]. Au cours des cinq dernières décennies, de nombreux algorithmes de reconnaissance de visages sophistiqués ont été proposés pour répondre à la demande des systèmes de reconnaissance de visage. Au début, la majorité des recherches en reconnaissance faciale a été axée sur l'utilisation des images d'intensité de visage (reconnaissance de visage 2D) en raison de leur simple acquisition en utilisant des capteurs à faible coût (appareils photo numériques).

Malgré les améliorations et la mise en œuvre de systèmes de reconnaissance de visage 2D qui ont un rendement acceptable dans des conditions contrôlées, des évaluations récentes font leur apparition comme FRVT (Face Recognition Vendor Test) en 2002. La recherche a montré [68] que la performance de ces systèmes se dégrade considérablement en fonction de pose, d'éclairage et des variations entre les instants de l'image de test et de galerie. Depuis le début des années 1990, sur la base des progrès constants dans la technologie des capteurs en trois dimensions, le processus d'acquisition 3D est devenu moins cher, plus rapide et plus fiable. En conséquence, de nombreux chercheurs se sont tournés vers les scans 3D du visage comme une solution potentielle pour les limites inhérentes de reconnaissance de visage 2D. Une enquête récente de ces tendances dans la communauté de reconnaissance de visage est donnée dans [69].

Dans ce chapitre, nous passons en revue la littérature de reconnaissance de visage 2D, 3D et multimodale 2D+3D. Nous commençons ce chapitre par un examen des importants algorithmes de reconnaissance de visage 2D, car la plupart des algorithmes de reconnaissance de visage 3D et 2D+3D sont des extensions directes d'algorithmes 2D existants. Le but de ce chapitre est d'évaluer les capacités et les lacunes des algorithmes en cours, afin d'identifier les problèmes et les orientations pour la recherche future dans ce domaine.

2.2 Reconnaissance de visage 2D

Jusqu'à maintenant, de nombreuses techniques ont été proposées pour la reconnaissance de visage à l'aide des images faciales 2D. Elles peuvent être généralement divisées en deux catégories principales: les approches globales et locales basées sur [70, 71]. Les approches globales utilisent l'ensemble du visage comme les données d'entrée du système de reconnaissance des visages, qui sont en outre projetées par rapport à un sous-espace de dimension relativement faible. On y trouve l'analyse de composantes principales (PCA) [72], l'analyse discriminante linéaire (LDA) [73], l'analyse en composantes indépendantes (ICA) [13], etc. Alors que celles basées sur les caractéristiques locales procèdent d'abord par localiser un certain nombre de caractéristiques ou composants de l'image du visage, puis les classer en combinant et en mesurant les statistiques locales correspondantes. On trouve dans ce cas: Elastic Bunch Graph Matching (EBGM) [74], Scale-Invariant Feature Transform (SIFT) [75] et LBP (Local Binary Patterns) [76]. Il existe aussi des méthodes qui n'appartiennent pas à l'une des deux classes, comme Actif Appearance Modèle (AAM) [77]. Dans les paragraphes suivants, nous présentons chaque catégorie en détails.

2.2.1 Approches globales

Les approches globales sont basées sur des techniques d'analyse statistique bien connues. Il n'est pas nécessaire de repérer certains points caractéristiques du visage comme les centres des yeux, les narines, le centre de la bouche, etc. sauf pour normaliser les images. Dans ces méthodes, les images de visage (qui peuvent être vues comme des matrices de valeurs de pixels) sont traitées de manière globale et sont généralement transformées en vecteurs, plus faciles à manipuler. L'avantage principal des méthodes globales est qu'elles sont relativement rapides à mettre en œuvre et que les calculs de base sont d'une complexité moyenne. En revanche, elles sont très sensibles aux variations

d'éclairement, de pose et d'expression faciale. Ceci se comprend aisément puisque la moindre variation des conditions de l'environnement ambiant entraîne des changements inéluctables dans les valeurs des pixels qui sont traités directement. Ces méthodes utilisent principalement une analyse de sous-espaces de visages. Cette expression repose sur un fait relativement simple: une classe de "formes" qui nous intéresse (dans notre cas, les visages) réside dans un sous-espace de l'espace de l'image d'entrée. Par exemple, considérons une petite image de 64×64 , en niveaux de gris codés sur 8 bits (donc de 0 à 255). Cette dernière possède 4096 pixels qui peuvent exprimer un grand nombre de classes de formes, comme des arbres, des maisons ou encore des visages. Cependant, parmi les 256^{4096} "configurations" possibles, seulement une petite quantité peut correspondre à des visages. Ainsi, la représentation de l'image originale est très redondante et la dimensionnalité de cette représentation pourrait être grandement réduite si l'on se concentre uniquement sur les formes qui nous intéressent. L'utilisation de techniques de modélisation de sous-espace a fait avancer la technologie de reconnaissance faciale de manière significative.

Nous pouvons distinguer deux types de techniques parmi les méthodes globales: les techniques linéaires et les techniques non linéaires [78].

2.2.1.1 Les techniques linéaires

Les techniques linéaires projettent linéairement les données d'un espace de grande dimension (par exemple, l'espace de l'image originale) sur un sous-espace de dimension inférieure. Malheureusement, ces techniques sont incapables de préserver les variations non convexes des variétés (géométriques au sens mathématique du terme) de visages afin de différencier des individus. Dans un sous-espace linéaire, les distances euclidiennes et plus généralement les distances de Mahalanobis, qui sont normalement utilisées pour comparer des vecteurs de données, ne permettent pas une bonne classification entre les classes de formes "visage" et "non-visage" et entre les individus eux-mêmes. Ce facteur crucial limite le pouvoir des techniques linéaires pour obtenir une détection et une reconnaissance du visage très précises.

La plus connue de ces approches est la technique dite des Eigenfaces présentée par Turk et Pentland dans [79]. L'analyse en composantes principales (ACP) est une méthode de réduction de dimension linéaire introduite par Pearson [80]. L'idée principale de l'ACP est de trouver une projection des données $\{x_i\}_{i=1}^n \in R^d$ qui maximise la variance. La base de ce nouveau sous-espace est formée des m premiers vecteurs propres de la matrice de covariance estimée:

$$c = \frac{1}{n} \sum_i (x_i - \bar{x})(x_i - \bar{x})^T \quad (2.1)$$

Avec \bar{x} , le vecteur moyenne estimée.

Si on range les vecteurs propres dans une matrice W de taille $d \times m$, la projection u_i de chaque donnée x_i est calculée par:

$$u_i = W^T(x_i - \bar{x}) \quad (2.2)$$

Les visages propres obtenus pour les 12 plus grandes valeurs propres pour la base de données ORL sont représentés dans la figure 2.1 [81].



Figure 2.1: Visages propres (vecteurs propres) des 12 plus grandes valeurs propres sont présentés provenant de la division ATT de la base de données ORL [81].

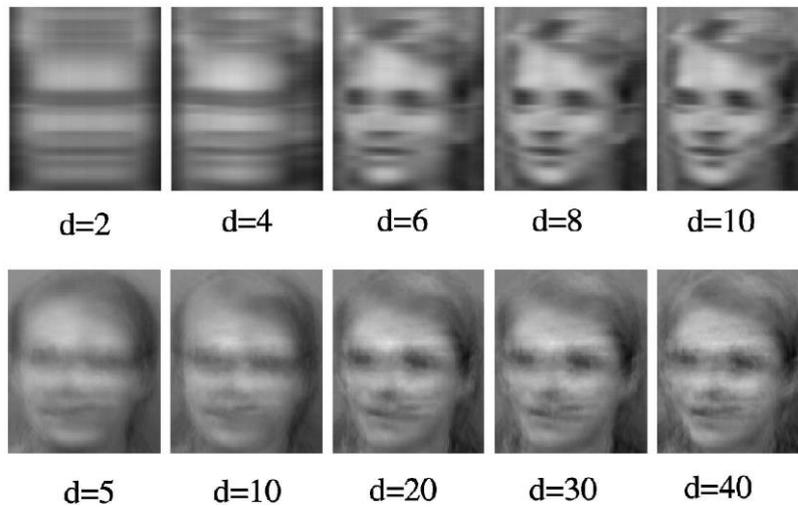


Figure 2.2: Certaines images reconstruites sur la base 2-D PCA (en haut) et PCA (en bas) sur la base de données ORL [82].

Le PCA classique nécessite que les images de visage soient mises sous formes de vecteurs, ce qui a pour effet de détruire la structure géométrique de l'image. Pour ne pas perdre les informations de voisinage lors du passage de l'image en vecteur, une méthode PCA bi-dimensionnelle (2-D PCA) [82] a été étudiée. Cette méthode prend en entrée des images et non plus des vecteurs. La figure 2.2 montre cinq images reconstruites d'une image de la base de données ORL en ajoutant au premier nombre des vecteur propres d ($d = 2, 4, 6, 8, 10$) des sous-images en même temps. Les images reconstruites apparaissent plus clairement que le nombre de sous-images est augmenté. A titre de comparaison, PCA (Eigenfaces) a également été utilisée pour représenter et reconstruire la même image de visage. La figure 4 montre également les images reconstruites lorsque le nombre de composantes principales d est fixé à 5, 10, 20, 30, et 40. L'APC n'a pas d'aussi bons résultats dans la reconstruction de cette image.

L'Analyse en composantes indépendantes (ou ICA pour Independent Component Analysis) a été introduite par Héroult, *Jutten* et *Ans* dans [83], [84]. Elle devint populaire lors de son utilisation dans le domaine du traitement du signal pour la séparation de sources aveugles dans [85] et [86]. Le but est d'exprimer un ensemble de n variables aléatoires x_1, \dots, x_n comme une combinaison linéaire de n variables aléatoires statistiquement indépendantes s_j :

$$x_j = a_{j,1}s_1 + \dots + a_{j,n}s_n \quad \forall j \tag{2.3}$$

ou sous forme matricielle: $x = As$.

Avec: $x = [x_1, \dots, x_n]^T$, $s = [xs_1, \dots, xs_n]^T$ et A est une matrice contenant les coefficients a_{ij} . Le but de l'analyse en composantes indépendantes est l'estimation des composantes originales s_i , ou de manière équivalente des coefficients a_{ij} . Par définition, les variables aléatoires s_i sont mutuellement indépendantes et la matrice de mélange est donc inversible. Ainsi le problème de l'ICA peut être formulé [87]:

$$u = Wx = WAs \tag{2.4}$$

Bartlett et al [89] montrent que les statistiques de premier et de second ordre détiennent des informations uniquement sur le spectre d'une image d'amplitude tout en jetant le spectre de phase. Cependant, certaines expériences ont mis en évidence que la capacité humaine à reconnaître des objets peut être due au spectre de phase. C'est aussi la raison pour laquelle ils ont étudié ICA comme un algorithme puissant pour la reconnaissance faciale. Depuis ICA sépare les moments d'ordre élevé de l'image d'entrée en plus des moments de second ordre utilisés par PCA. Cela peut être considéré comme une généralisation de PCA. En outre, ICA permet une meilleure caractérisation des données dans un espace à n dimensions et les vecteurs de base trouvés par ICA ne sont pas nécessairement orthogonaux afin qu'ils réduisent également l'erreur de reconstruction. *Bartlett et al* [89] ont fourni deux architectures différentes pour l'ICA: une première architecture (ICA I) qui construit une base d'images statistiquement indépendantes et une deuxième architecture (ICA II) qui fournit une représentation en code factoriel des données. La figure 2.3 présente l'ICA en code factoriel pour la base de données ORL. Pour plus de détails sur la théorie et les applications possibles de l'analyse en composantes indépendantes, on peut se référer à [88]. Pour l'application de l'ICA à la reconnaissance de visages, on peut consulter [89] et [90].



Figure 2.3: Vecteurs de base ICA provenant de la division ATT de la base de données ORL [81].

Le Multidimensional Scaling (MDS) est une autre méthode très connue de réduction de dimension linéaire [91]. Au lieu de conserver la variance des données au cours de la projection, elle s'efforce de préserver toutes les distances entre chaque paire d'exemples $dist(x_i, x_j)$ en cherchant une transformation linéaire qui minimise l'énergie:

$$\varepsilon_{m ds} = \sum_{i,j=1}^n (dist(x_i, x_j) - \|u_i - u_j\|)^2 \tag{2.5}$$

Ce problème de minimisation peut être résolu par une décomposition en valeurs propres [92], [93]. Lorsque la distance entre les données est la distance euclidienne, les sorties $u_i \in R^m$ de la MDS sont les mêmes que les sorties de PCA. Elles sont obtenues par une rotation suivie d'une projection sur les axes de plus grande variance.

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) [94]. L'algorithme de la NMF, comme celui de PCA, 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 dans les poids de la combinaison. Au final, certains vecteurs de l'espace réduit par PCA, (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.

Il existe d'autres techniques également construites à partir de décompositions linéaires comme l'analyse discriminante linéaire (LDA) [95]. Tandis que la PCA construit un sous-espace pour représenter de manière "optimale" seulement "l'objet" visage, la LDA construit un sous-espace discriminant pour distinguer de façon "optimale" les visages de différentes personnes. Appelée également "Fisher Linear Discriminant", la LDA 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 qui date de 1936 [96]. Cette méthode fut appliquée aux visages en 1996 [97]. Elle comporte deux étapes: une réduction de l'espace d'origine par le PCA, puis les vecteurs de l'espace de projection final, appelés "Fisher faces", 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 total de l'approche LDA (pour plus de détails, voir le chapitre 3). Des études comparatives (par exemple [95]) montrent que les méthodes basées sur la LDA donnent généralement de meilleurs résultats que les méthodes basées sur la PCA. Les visages propres obtenus pour les 12 plus grandes valeurs propres de LDA pour la base de données ORL sont représentés sur la figure 2.4.

Beaucoup d'efforts ont été faits pour améliorer les techniques linéaires d'analyse de sous-espace depuis que PCA, ICA et LDA ont été proposées pour la reconnaissance faciale. Par exemple, *Pentland et al.* [98] ont modifié la PCA aux visages pour faire face aux variations de pose. Le sous-espace probabiliste [99] a été introduit pour obtenir une mesure de similarité plus significative dans le cadre probabiliste. *Lu et al.* [100] ont présenté une hybridation entre la D-LDA (LDA direct) et la F-LDA (Fractional LDA), une variante de la LDA, dans laquelle les fonctions pondérées sont utilisées pour éviter l'erreur de classification causée par catégories de produits qui sont trop près.

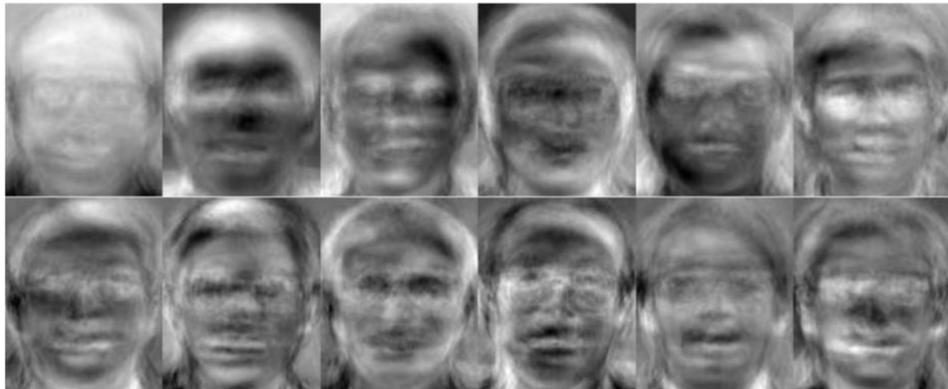


Figure 2.4: Premiers 12 vecteurs LDA provenant de la division ATT de la base de données ORL [81].

Vasilescu et Terzopoulos [101] ont proposé une approche basée sur le tenseur multilinéaire de décomposition des ensembles d'images pour résoudre la confusion de plusieurs facteurs limités dans le même système de reconnaissance de visage, tels que l'éclairage et la pose. Il y a d'autres techniques linéaires qui ont également été utilisées pour le calcul de vecteurs caractéristiques:

- Regularized Discriminant Analysis (RDA) [102].
- Regression LDA (RLDA) [103].
- Null-space LDA (NLDA) [104] [105].
- Direct-weighted LDA [105].
- Dual-space LDA [106].
- Pair-wise LDA [107].
- Generalized Singular Value Decomposition [108] [109].

- Direct Fractional-Step LDA [100].
- Boosting LDA [110].
- Discriminant Local Feature Analysis [111].
- Block LDA [112].
- Enhanced FLD [113].
- Incremental LDA [114].
- Discriminants Vecteurs Communs (DCV) [115].
- L'analyse discriminante bilinéaire (BDA) [116],
- La technique dite de « Vecteurs communs discriminants » (DCV) [115].

Bien que ces méthodes globales linéaires basées sur l'apparence évitent l'instabilité des toutes premières méthodes géométriques qui ont été mises au point, elles ne sont pas assez précises pour décrire les subtilités des variétés (géométriques) présentes dans l'espace de l'image originale. Ceci est dû à leurs limitations à gérer la non-linéarité en reconnaissance faciale: les déformations de variétés non linéaires peuvent être lissées et les concavités peuvent être remplies, causant des conséquences défavorables.

2.2.1.2 Les techniques non-linéaires

Les méthodes précédentes trouvent un espace de représentation fidèle aux données lorsque la structure de ces données est linéaire ce qui n'est généralement pas le cas. L'idée est d'utiliser une fonction noyau pour construire un espace de grande dimension dans lequel le problème devient linéaire. On peut ainsi appliquer des méthodes linéaires de réduction de dimension lorsque la structure intrinsèque des données n'est pas linéaire. Ces méthodes utilisent généralement « l'astuce du noyau » (de l'anglais kernel trick) qui établit que tout algorithme formulé avec une fonction noyau peut être reformulé avec une autre fonction noyau. Une démarche courante est d'exprimer la méthode avec un produit scalaire, puis de le remplacer par une fonction noyau. Le kernel trick permet alors de travailler dans l'espace transformé sans avoir à calculer explicitement l'image de chaque donnée. L'Analyse en Composantes Principales à Noyaux (Kernel Principal Component Analysis, KPCA) est la reformulation non linéaire de la technique linéaire classique qu'est l'analyse en composantes principales en utilisant des fonctions à noyaux [117]. Depuis plusieurs années, la reformulation de techniques classiques à l'aide de « l'astuce du noyau » a permis l'émergence de nombreuses techniques comme les machines à support de vecteurs (Support Vector Machine, SVM) [118]. PCA à noyaux calcule les principaux vecteurs propres de la matrice de noyaux plutôt que la matrice de covariance. Cette reformulation de PCA classique peut être vue comme une réalisation de PCA sur l'espace de grande dimension transformée par la fonction noyau associée. PCA à noyaux permet ainsi de construire des mappings non linéaires. PCA à noyaux calcule d'abord la matrice de noyaux K des points x_i dont les entrées sont définies par [118]:

$$k_{i,j} = k(x_i, x_j) \tag{2.6}$$

Où k est la fonction noyau. Ensuite, la matrice de noyaux K est centrée :

$$k_{i,j} = k_{i,j} - \frac{1}{n} \sum_l k_{i,l} - \frac{1}{n} \sum_l k_{j,l} + \frac{1}{n^2} \sum_{lm} k_{lm} \tag{2.7}$$

Cette opération correspond à la soustraction de la moyenne des vecteurs caractéristiques dans le PCA linéaire classique. Les d principaux vecteurs propres v_i de la matrice de noyaux centrée sont ensuite calculés. Il peut être montré que les vecteurs propres α_i de la matrice de covariance (dans l'espace de grande dimension) sont des versions mises à l'échelle des vecteurs propres v_i de la matrice de noyaux:

$$\alpha_i = \frac{1}{\sqrt{\lambda_i}} v_i \quad (2.8)$$

La technique KPCA est une méthode basée sur les noyaux et ses performances dépendent alors grandement du choix de la fonction noyau k . Les noyaux classiquement utilisés sont le noyau linéaire (cela revient alors à effectuer la PCA classique), le noyau polynomial ou encore le noyau gaussien [118]. L'analyse en composantes principales à noyaux a été appliquée avec succès à plusieurs problèmes comme la reconnaissance de la parole [119], ou la détection de nouveaux éléments d'un ensemble [120]. Un gros défaut de l'analyse en composantes principales à noyau est que la taille de la matrice de noyaux est le carré du nombre d'échantillons de l'ensemble d'apprentissage ce qui peut rapidement être prohibitif.

D'autres techniques non linéaires ont également été utilisées dans le contexte de la reconnaissance faciale :

- Le KICA (Kernel Independent Component Analysis) [121].
- Le Maximum Variance Unfolding (MVU) [122].
- l'Isomap dans [123].
- Les diffusions maps dans [124].
- Le Local Linear Embedding (LLE) dans [125] ou [126].
- Locality Preserving Projection (LPP) [127].
- Embedded Manifold [128].
- Nearest Manifold Approach [129].
- Discriminant Manifold Learning [130].
- Les Laplacian eigenmaps dans [131], [132] ou [133].
- Le Hessian LLE [134].
- Le Local Tangent Space Analysis (LTSA) dans [135].
- Les approches neuronales dans [133] ou [136] (auto encodeurs), dans [137] (cartes de Kohonen) et dans [138] (réseaux de neurones convolutionnels).
- Exponential Discriminant Analysis (EDA) [139].

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

2.2.2 Approches locales

Les méthodes locales peuvent être classées en deux catégories, les méthodes basées sur les points d'intérêt et celles basées sur l'apparence du visage. Dans le premier cas, on détecte tout d'abord les points d'intérêt et ensuite on extrait des caractéristiques localisées sur ces points d'intérêt. Dans le second cas, on divise le visage en petites régions (ou patches) sur lesquelles les caractéristiques locales sont directement extraites. En comparaison avec les approches globales, les méthodes locales présentent certains avantages. Tout d'abord, les approches locales peuvent fournir des informations supplémentaires basées sur les parties locales. De plus, pour chaque type de caractéristiques locales, on peut choisir le classificateur le plus adapté.

Malgré ces avantages, l'intégration d'informations de structure plus globale est nécessaire. En général, il y a deux façons de procéder pour atteindre cet objectif. Premièrement, les informations globales sont intégrées dans les algorithmes en utilisant des structures de données, telles qu'un graphe où chaque nœud représente une caractéristique locale alors qu'une arête entre deux nœuds représente la relation spatiale entre eux. La reconnaissance de visage apparaît comme un problème d'appariement de deux graphes. Deuxièmement, les algorithmes peuvent utiliser des techniques de fusion de scores:

des classificateurs séparés sont utilisés sur chaque caractéristique locale pour calculer une similarité et ensuite les similarités obtenues sont combinées afin d'obtenir un score global pour la décision finale [140].

2.2.2.1 Méthodes locales basées sur les caractéristiques d'intérêt

Les méthodes les plus anciennes en reconnaissance de visages appartiennent à cette catégorie [141, 142, 143, 144, 145]. Elles s'appuient toutes [141, 143] sur l'extraction de caractéristiques géométriques spécifiques telles que la largeur de la tête, les distances entre les yeux, etc. Ces données sont ensuite utilisées par des classificateurs afin de reconnaître des individus. Ces méthodes présentent les deux inconvénients suivants: (1) les caractéristiques géométriques sont difficiles à extraire dans certains cas puisque la tâche de détection précise de points caractéristiques n'est pas facile, en particulier dans les cas où des occultations ou des variations (pose, expression) de visages sont présentes et (2) les caractéristiques géométriques seules ne sont pas suffisantes pour représenter entièrement un visage et d'autres informations utiles telles que les valeurs des niveaux de gris de l'image sont complètement écartées.

Ces deux limites ont engendré deux directions de recherche. La première se concentre sur les performances des détecteurs de points caractéristiques du visage. Dans [143], *Brunelli et Poggio* ont proposé d'utiliser un ensemble d'apprentissage pour détecter la position de l'œil dans une image. Ils ont tout d'abord calculé pour chaque point des coefficients de corrélation entre l'image de test et les images de l'ensemble d'apprentissage et ensuite ils ont cherché les valeurs maximales. *Rowley et al.* [144] ont utilisé plusieurs détecteurs de traits spécifiques correspondant à chaque partie du visage, telles que les yeux, le nez, la bouche, etc. *Lanitis et al.* [145] ont proposé de construire des modèles statistiques de la forme du visage. Malgré toutes ces recherches, il n'existe pas encore de détecteur de points caractéristiques qui soit suffisamment fiable et précis. Dans la deuxième direction, les méthodes se concentrent sur des représentations plus élaborées des informations portées par les points caractéristiques du visage, plutôt que simplement sur des caractéristiques géométriques. *Manjunath et al.* [142] ont proposé des algorithmes pour détecter et représenter des caractéristiques faciales à partir d'ondelettes de *Gabor*. Pour chaque point détecté, deux types d'information sont stockées: sa position et ses caractéristiques (les caractéristiques sont extraites en utilisant le filtre de *Gabor* sur le point considéré). Pour modéliser la relation entre les points caractéristiques, un graphe topologique est construit pour chaque visage. Basée sur cette idée, *Lades et al.* [146] ont proposé d'utiliser un graphe topologique déformable au lieu d'un graphe topologique fixe comme dans [142] afin de proposer un modèle de représentation du visage appelé Dynamic Link Architecture (DLA). Cette approche permet de faire varier le graphe en échelle et en position en fonction des variations d'apparence du visage considéré. Le graphe est une grille rectangulaire localisée sur l'image (voir figure 2.4) où les nœuds sont étiquetés avec les réponses des filtres de *Gabor* dans plusieurs orientations et plusieurs fréquences spatiales appelées jets. Les bords sont étiquetés par des distances, où chaque bord relie deux nœuds sur le graphe. La comparaison entre deux graphes de visage est réalisée en déformant et en mettant en correspondance le graphe représentatif de l'image de test avec chacun des graphes représentatif des images de référence.

Plus tard, *Wiskott et al.* [74] ont étendu l'utilisation de LDA à une méthode très connue appelée Elastic Bunch Graph Matching (EBGM), où les nœuds des graphes sont situés sur un certain nombre de points sélectionnés du visage. De manière similaire à la méthode de [142], *Wiskott et al.* ont utilisé les ondelettes de *Gabor* pour extraire les caractéristiques des points détectés car les filtres de *Gabor* sont robustes aux changements d'illumination, aux distorsions et aux variations d'échelle (voir les détails sur les filtres de *Gabor* dans le chapitre 3). De ce fait, l'EBGM fut l'un des algorithmes les plus performants lors de la compétition de FERET en 1996. Le succès de ces méthodes a aussi été une motivation pour certains travaux plus récents [147, 148].

Pour conclure, de nombreuses méthodes basées sur l'extraction de points caractéristiques ont été proposées et ces méthodes peuvent être utiles et efficaces pour la reconnaissance de visages dans le cas

où une seule image de référence est disponible. Cependant, leurs performances dépendent beaucoup de l'efficacité des algorithmes de localisation des points caractéristiques du visage. En pratique, la tâche de détection précise des points caractéristiques n'est pas facile et n'a pas été résolue de manière fiable, en particulier dans les cas où la forme ou l'apparence d'une image du visage peuvent beaucoup varier [140].

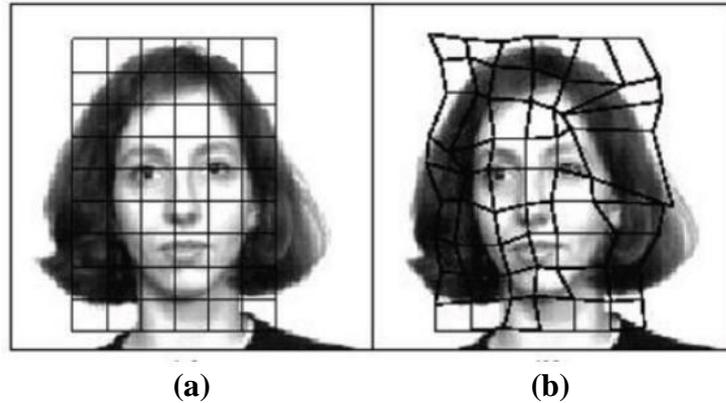


Figure 2.5: Exemple de grille d'appariement. (a) grille de référence, (b) grille correspondante.

2.2.2.2 Les méthodes locales basées sur l'apparence du visage

Une fois les régions locales définies, il s'agit de choisir la meilleure manière de représenter les informations de chaque région. Cette étape est critique pour les performances du système de reconnaissance. Les caractéristiques couramment utilisées sont: les coefficients de *Gabor* [74, 142, 143], les ondelettes de *Harr* [149], les transformées de Fourier, SIFT (Scale Invariant Feature Transform) [150], les caractéristiques basées sur la méthode LBP (Local Binary Pattern) [151], LPQ (Local Phase Quantization) [152], WLD (Weber Law Descriptor) [153], BSIF (Binarized Statistical Image Features) [154]. Dans cette section on discutera en détails la méthode LBP et ses variantes.

- **LBP et sa variante récente [155]**

La méthode LBP originale étiquette les pixels d'une image avec des nombres décimaux, qui sont appelés LBP qui codent la structure locale autour de chaque pixel. On procède ainsi, comme illustré sur la figure 2.6: chaque pixel est comparé à ses huit voisins dans un voisinage 3×3 en soustrayant la valeur du pixel central; les valeurs strictement négatives résultantes sont codées avec 0 et les autres avec un. Pour chaque pixel donné, un nombre binaire est obtenu en concaténant l'ensemble de ces valeurs binaires dans le sens horaire, ce qui commence à partir de l'une de ses voisines en haut à gauche. La valeur décimale correspondant du nombre binaire généré est ensuite utilisée pour marquer le pixel donné. Les nombres binaires dérivés sont appelés codes de LBP [156]. Pour plus de détails, on peut se référer au chapitre 4.

La méthodologie de LBP a été développée récemment avec un grand nombre de variations pour l'amélioration des performances dans différentes applications. Ces variations portent sur différents aspects de l'opérateur LBP original: 1) l'amélioration de sa capacité discriminatoire; 2) l'amélioration de sa robustesse; 3) la sélection de son voisinages. Dans cette section, nous passons en revue les dernières variantes de LBP (voir tableau 2.1 pour la liste).

1) *Renforcer la capacité discriminative*

La méthode LBP définit un certain nombre de modèles pour décrire les structures locales. Pour améliorer leur capacité discriminante, plus de modèles ou d'informations peuvent être codés. Jin et al. [157] ont amélioré l'opérateur de LBP à décrire plus d'informations de structure locale dans certaines

circonstances. Plus précisément, ils ont proposé un LBP amélioré (Improved LBP, ILBP), qui compare tous les pixels avec l'intensité moyenne de tous les pixels (y compris le pixel central) du bloc (comme le montre la figure 2.6). Par exemple, l'opérateur LBP (8,1) ne produit que 256 (2^8) modèles dans un voisinage 3×3 , tandis que ILBP à 511 modèles ($2^9 - 1$). Plus tard, ILBP a été étendue pour utiliser toutes les tailles à la place du bloc original 3×3 [158]. Presque au même moment, un système similaire a été utilisé pour étendre CT à CT modifiée [159] à savoir MLBP (Modified LBP) [160]. Un opérateur LBP moyen est présenté dans [161], qui est similaire à ILBP, mais sans tenir compte des pixels centraux.

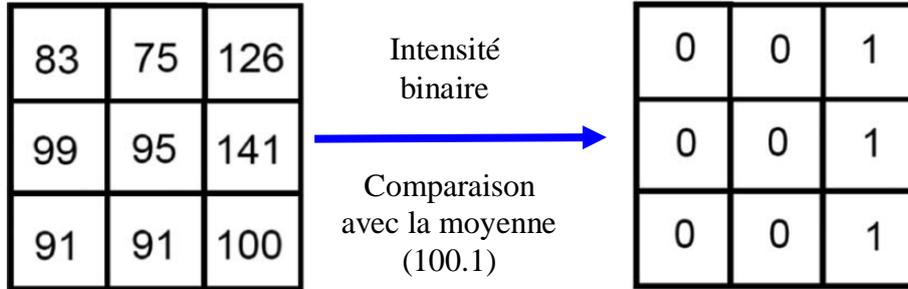


Figure 2.6: Exemple de la méthode ILBP [162]

Yang et Wang [163] ont proposé Hamming LBP pour améliorer la capacité discriminative de la LBP originale. Ils ont reclassé les motifs non uniformes sur la base de la distance de Hamming, au lieu de les rassembler en un seul bin comme fait LBP^{u2}.

ELBP [164, 165] est une autre approche pour améliorer la capacité discriminative de LBP. L'opérateur ELBP effectue non seulement la comparaison binaire entre le pixel central et ses voisins, mais encode aussi leurs différences de valeurs de gris (Gray-value Differences, GDs) en utilisant des unités binaires. Plus précisément, la fonction ELBP se compose de plusieurs codes de LBP à multiples couches, qui codent le GD entre le pixel central et ses pixels voisins. Comme le montre la figure 2.7, la première couche de ELBP est en fait le code de LBP originale qui code le signe de la GD. Les couches suivantes de ELBP codent la valeur absolue de GD. Fondamentalement, chaque valeur absolue GD est d'abord codée dans sa représentation binaire, puis toutes les valeurs binaires de chaque couche donnent un LBP supplémentaire. Par exemple, dans la figure 2.7, la première couche est le code de LBP originale qui code le signe de la GD, donnant ainsi un nombre décimal 211 à sa forme binaire $(11010011)_2$. Les valeurs absolues des GD, c'est à dire, 1, 5, 3, 2, 1, 2, 3, et 0 sont d'abord codées dans leurs formes binaires: $(001)_2$, $(101)_2$, $(011)_2$, $(010)_2$, . . ., etc. Utilisant un même schéma de poids de LBP sur tous les bits binaires, son code ELBP de la couche correspondante peut être généré. Par exemple, L2 est composé de $(01000000)_2$ et sa valeur décimale est 64; L3 est composé de $(00110110)_2$ et sa valeur décimale est 54. Enfin, L4 est composé de $(11101010)_2$ et sa valeur décimale est 234. En conséquence, lors de la description de textures locales similaires, bien que le premier LBP de couche n'est pas assez discriminant, l'information codée dans les autres couches supplémentaires peuvent être utilisées pour les distinguer. Son inconvénient est que ELBP augmente la fonction dimensionnalité dans une large mesure.

Plus récemment, Guo et al. [166] ont proposé CLBP (Complete LBP), qui, à notre avis, est tout à fait similaire avec ELBP. En outre, la lombalgie chronique comprend à la fois le signe et le GDs entre un pixel central donné et ses voisins afin d'améliorer le pouvoir discriminant de l'opérateur LBP original. Contrairement à la stratégie de codage binaire utilisé par ELBP, CLBP compare la valeur absolue de GD avec le pixel central donné pour ajouter un nouveau code de LBP.

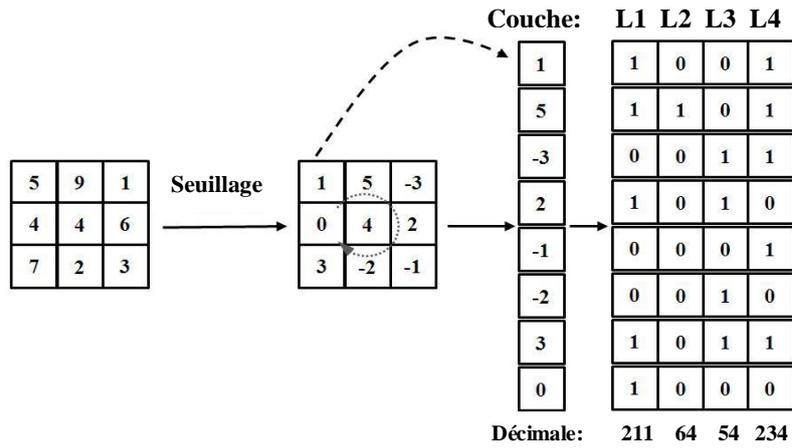


Figure 2.7: Exemple de la méthode ELBP [164]

2) L'amélioration de la robustesse

LBP est sensible au bruit, puisque l'opérateur seuille exactement à la valeur de pixel central. Pour résoudre ce problème, Tan et Triggs [167] ont étendu LBP original pour une version avec 3 codes de valeurs, appelé LTP (Local Ternary Patterns). Dans LTP, indicateur signe de LBP est remplacé par:

$$s(i_n, i_c, t) = \begin{cases} 1, & i_n \geq i_c + t \\ 0, & |i_n - i_c| < t \\ -1, & i_n \leq i_c - t \end{cases} \quad (2.9)$$

Où t est égal à un seuil défini par l'utilisateur. Les codes LTP sont plus résistants aux bruits, mais ne sont plus strictement invariants aux transformations au niveau gris. Un schéma de codage est utilisé pour diviser chaque motif ternaire en deux parties: l'une positive et l'autre négative, comme illustré sur la figure 2.8. Un problème de LTP est de régler le seuil t , ce qui n'est pas simple.

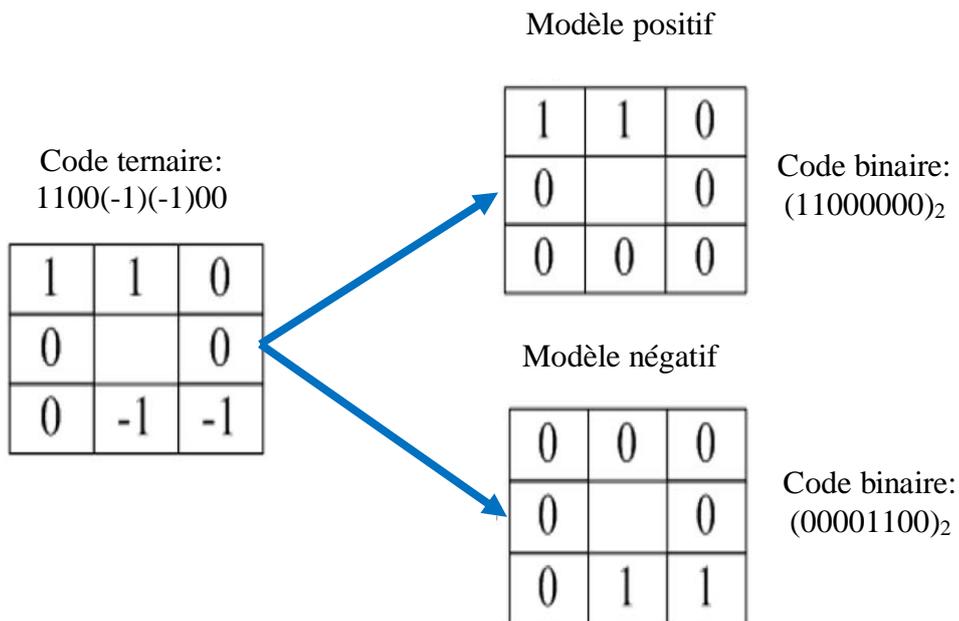


Figure 2.8: Exemple de la méthode LTP [167].

SLBP (Soft LBP) a été introduite dans [168] qui emploie deux fonctions d'appartenance floues au lieu de l'indicateur du signe de LBP:

$$s_{1,d}(x) = \begin{cases} 0, & x < -d \\ 0.5 + 0.5 \frac{x}{d}, & -d \leq x \leq d \\ 1, & x > d \end{cases} \quad (2.10)$$

$$s_{0,d}(x) = 1 - s_{1,d}(x) \quad (2.11)$$

d paramètre contrôlant la quantité de fuzzification, qui est effectuée par la fonction floue. Lorsque le voisinage se compose de points de prélèvement de p . SLBP améliore la robustesse dans le sens d'un petit changement de l'image d'entrée et ne provoque qu'une faible variation de la sortie. Cependant, il perd de l'invariance de variations monotones et accroît ainsi la complexité de calcul. Comme pour la LTP, une bonne valeur de d doit être réglée.

3) Choix des voisinages

Le choix d'un voisinage approprié pour les techniques basées sur LBP a un impact significatif sur la performance finale. Elle implique le nombre de points d'échantillonnage, la répartition des points d'échantillonnage, la forme et la taille du voisinage. Les pixels voisins de LBP originale sont définis sur un cercle. *Liao et Chung* [169] ont fait valoir que la raison principale pour définir les voisinages d'une façon isotrope est d'obtenir l'invariance par rotation pour la description de la texture. Cependant, cela n'est pas adapté à toutes les applications. Au contraire, l'information anisotrope peut également être une caractéristique importante. En conséquence, ils ont proposé allonger LBP avec des pixels voisins se trouvant sur une ellipse. La figure 2.9 montre deux exemples de la LBP allongée, où A et B désignent le grand axe et le petit axe respectivement et m est le nombre de pixels voisins. Suivi par la LBP originale, la technique d'interpolation bilinéaire est adoptée pour les pixels voisins qui ne sont pas exactement au niveau des pixels. L'opérateur de LBP allongée peut être tourné autour du pixel central, avec un angle spécifique pour caractériser les structures locales allongées dans des orientations différentes, pour réaliser l'analyse de multi orientation.

Afin de saisir non seulement les microstructures, mais aussi les macrostructures, *Li et al.* [171], [171] ont proposé une LBP multibloc (multiblock LBP, MB-LBP), qui, au lieu de comparer les pixels, compare les intensités moyennes des sous-régions voisines. LBP originale peut être considérée comme un cas particulier de la MB-LBP. La figure 2.10 montre un exemple de MB-LBP, où chaque sous-région est composée de six pixels.

La sous-région peut être soit un rectangle ou un carré. Un système similaire est introduit dans [172]: TP-LBP (Three-patch) et FP-LBP (Four-patch LBP) sont proposées pour comparer les distances entre les blocs au lieu des pixels [156] ou intensité moyenne [171] et une fonction de distance peuvent être utilisée (par exemple, norme L2 de leurs différences de niveau de gris).

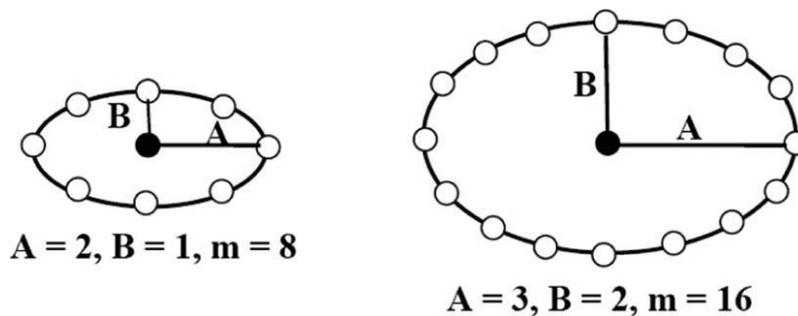


Figure 2.9: Exemple de la méthode LBP allongée [156].

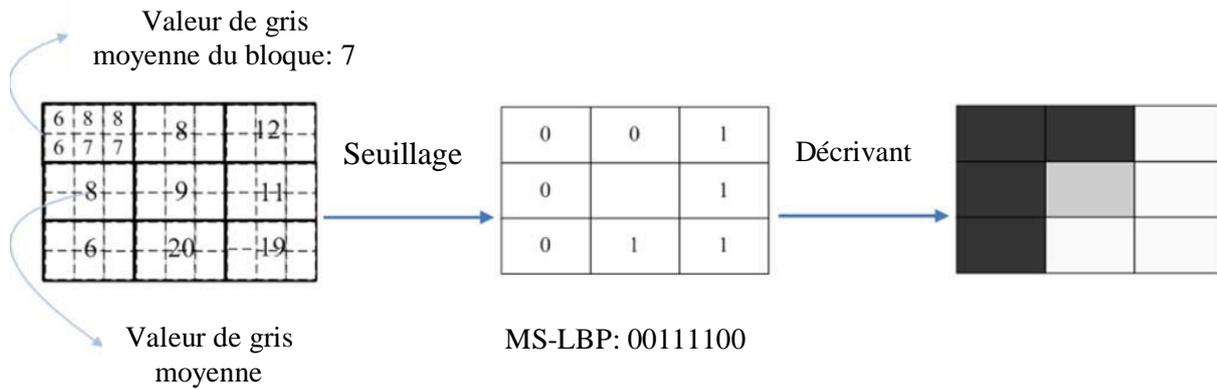


Figure 2.10: Exemple de la méthode MB-LBP [170].

Tableau 2.1: Liste des variantes récentes de la LBP

Subdivision	Variante	Propriétés	Référence
1) Renforcer la capacité discriminative	Improved LBP	Examen des effets des pixels centraux; présentation des modèles de structure complète	[157, 160, 161]
	Hamming LBP	Intégration des modèles non-uniformes dans les modèles uniformes	[163]
	Extended LBP	Différenciation des mêmes motifs binaires locaux; induction d'une forte dimension	[164, 165]
	Completed LBP	Prise en compte à la fois du signe et de l'information de grandeur de la région locale donnée	[166]
2) L'amélioration de la robustesse	Local Ternary Patterns	Apport d'un nouveau seuil plus strictement invariant pour la transformation de niveau gris	[167]
	Soft LBP	Pas invariant aux changements de niveaux de gris monotones; cause la complexité de calcul	[167]
3) Choisir les voisinages	Elongated LBP	Extraction de l'information anisotrope et de perte des informations anisotropes; pas invariant à la rotation	[169]
	Multi-Block LBP	Capture micro et macro informations de structure	[170, 171]
	Three/Four Patch LBP	Encodage de l'information de patch	[172]

2.2.3 Autres approches

D'autres approches basées sur des modèles statistiques ont également été largement explorées. Un modèle statistique est une formalisation des relations entre les variables sous la forme d'équations mathématiques qui décrivent comment une ou plusieurs variables aléatoires sont liées à une ou plusieurs variables aléatoires. Le modèle est considéré comme statistique lorsque les variables ne sont pas déterministes mais stochastiquement liées. Pour formuler l'analyse du visage avec le modèle statistique, nous allons construire des modèles de l'apparence du visage et de ses variations et apprendre les manières dont la forme et la texture du visage humain varient dans une série d'images. Des exemples des approches basées sur des modèles statistiques comprennent le modèle actif d'apparence (Active Appearance Model, AAM) [173], GMM (Gaussian Mixture Model) [174,175], 1D HMM (Hidden Markov Model) [176], pseudo-2D HMM [177], etc.

Dans cette sous-section, un représentant typique du modèle basé sur une technique statistique AAM est introduit en détail non seulement en raison de sa bonne application dans la reconnaissance de visage, mais en raison de ses contributions fondamentales à l'analyse de l'image du visage.

2.2.3.1 Modèle actif d'apparence (AAM)

Il a déjà été démontré [178] que le modèle actif d'apparence est un outil puissant permettant de représenter des visages de façon réaliste et naturelle. Il se base sur une technique PCA permettant de représenter conjointement les variations de forme et de texture présentes dans un ensemble d'apprentissage. En effet, après avoir aligné toutes les formes de l'ensemble d'apprentissage par rapport à la forme moyenne à l'aide d'une transformation Procrustéenne [179], le modèle statistique de forme est donné par :

$$s_i = \bar{s} + \Phi_s b_{si} \quad (2.12)$$

Où s_i est la forme synthétisée, Φ_s est une matrice contenant les principaux modes de variation de forme et b_{si} est un vecteur contrôlant la forme synthétisée. Il est alors possible de déformer les textures de l'ensemble d'apprentissage sur la forme moyenne \bar{s} pour séparer la texture de la forme. De manière similaire, après avoir calculé la texture sous forme moyenne \bar{g} et normalisé toutes les textures de l'ensemble d'apprentissage par rapport à la texture moyenne à l'aide d'une translation et d'une remise à l'échelle des niveaux de gris, le modèle statistique de texture est donné par :

$$g_i = \bar{g} + \Phi_t b_{ti} \quad (2.13)$$

Où g_i est la texture sous forme synthétisée, Φ_t est une matrice contenant les modes principaux de variation de texture dans l'ensemble d'apprentissage et b_{ti} est un vecteur contrôlant la texture synthétisée, sans forme. Après pondération des valeurs des composantes de b_{si} avec des poids convenablement choisis pour rendre leur ordre de grandeur comparable à b_{ti} [178], ces vecteurs sont concaténés. En pratique les composantes de b_{si} sont pondérées par un poids égal à la racine carrée du rapport de la somme des valeurs propres associées à Φ_t sur la somme des valeurs propres associées à Φ_s . Une PCA supplémentaire sur les nouveaux vecteurs obtenus retourne le modèle statistique d'apparence donné par :

$$s_i = \bar{s} + Q_s c_i \quad (2.14)$$

$$g_i = \bar{g} + Q_t c_i \quad (2.15)$$

Où: Q_s et Q_t sont des matrices contenant les principaux modes de variation d'apparence dans l'ensemble d'apprentissage et c_i est un vecteur de paramètres d'apparence permettant de contrôler simultanément la forme et la texture de l'objet. Une autre manière de représenter le modèle combiné consiste à concaténer directement, pour chaque image i de l'ensemble d'apprentissage, les vecteurs de forme s_i et de texture g_i pour former le vecteur b_{sti} . Les vecteurs s_i seront multipliés par une matrice de poids W_s convenablement choisis afin de rendre leur ordre de grandeur comparable à g_i . En pratique, les vecteurs s_i sont pondérés par un poids égal au rapport de l'écart-type des intensités normalisées par l'écart-type des formes normalisées de l'ensemble d'apprentissage. En appliquant une PCA aux vecteurs b_{sti} obtenus, le modèle d'apparence modifié est donné par:

$$\begin{pmatrix} W_s s_i \\ g_i \end{pmatrix} = \begin{pmatrix} \bar{s} \\ \bar{g} \end{pmatrix} + Q_{st} c_i \quad (2.16)$$

Où Q_{st} est une matrice représentant les principaux modes de variation des vecteurs combinés de forme et de texture et c_i est le vecteur de paramètres d'apparence contrôlant conjointement la forme et la texture synthétisées par le modèle.

Ayant le vecteur de paramètres c_i , la forme et la texture correspondantes peuvent être retrouvées respectivement à partir des équations (2.14) et (2.15) pour la représentation standard et (2.16) pour la

représentation modifiée. La texture sans forme reconstruite est alors projetée sur la forme reconstruite afin d'obtenir l'apparence globale du visage reconstruit. De plus, il est nécessaire de rajouter au vecteur de paramètres d'apparence c_i , un vecteur de paramètres de pose p_i permettant de contrôler l'échelle, l'orientation et la position du visage reconstruit. En partant du principe que les vecteurs d'apparence c et de pose p constituent une représentation d'un visage, le modèle actif d'apparence AAM est capable d'ajuster automatiquement ces paramètres à un visage cible inconnu [180] en minimisant une image résiduelle $r(c, p)$ qui n'est autre que la différence de texture entre le visage synthétisé et le masque de l'image cible correspondant. Ces résidus sont utilisés pour calculer les matrices R_a et R_t permettant de relier linéairement les paramètres d'apparence et de pose aux résidus correspondants $\delta(c) = -R_a r(c, p)$ et $\delta(p) = -R_t r(c, p)$, de telle sorte à minimiser $|r((c, p) + \delta(c, p))|^2$. Un développement de Taylor du premier ordre donne la solution suivante:

$$R_a = \left(\frac{\partial r^T}{\partial c} \frac{\partial r}{\partial c} \right)^{-1} \frac{\partial r^T}{\partial c} \quad \text{et} \quad R_t = \left(\frac{\partial r^T}{\partial p} \frac{\partial r}{\partial p} \right)^{-1} \frac{\partial r^T}{\partial p} \quad (2.17)$$

Avec le terme $\frac{\partial r}{\partial p}$ constant estimé à partir de l'ensemble d'apprentissage en déplaçant les paramètres d'apparence et de pose de leurs valeurs optimales. Les paramètres d'apparence et de pose sont alors adaptés au visage cible à l'aide d'une procédure itérative d'optimisation. L'adaptation du modèle standard sur un visage inconnu en partant de la forme et de la texture moyenne est illustrée sur la figure 2.11.

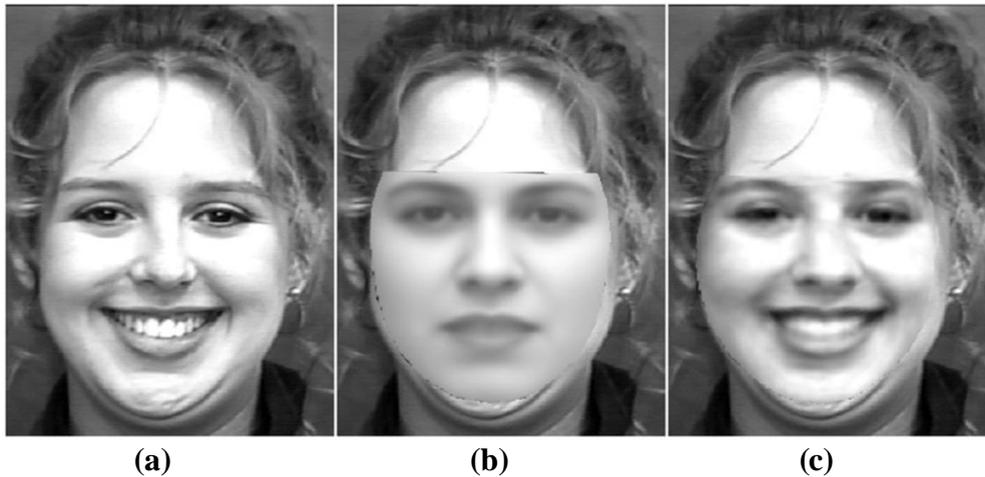


Figure 2.11: (a) : visage inconnu, (b): initialisation du modèle et (c): adaptation du modèle classique

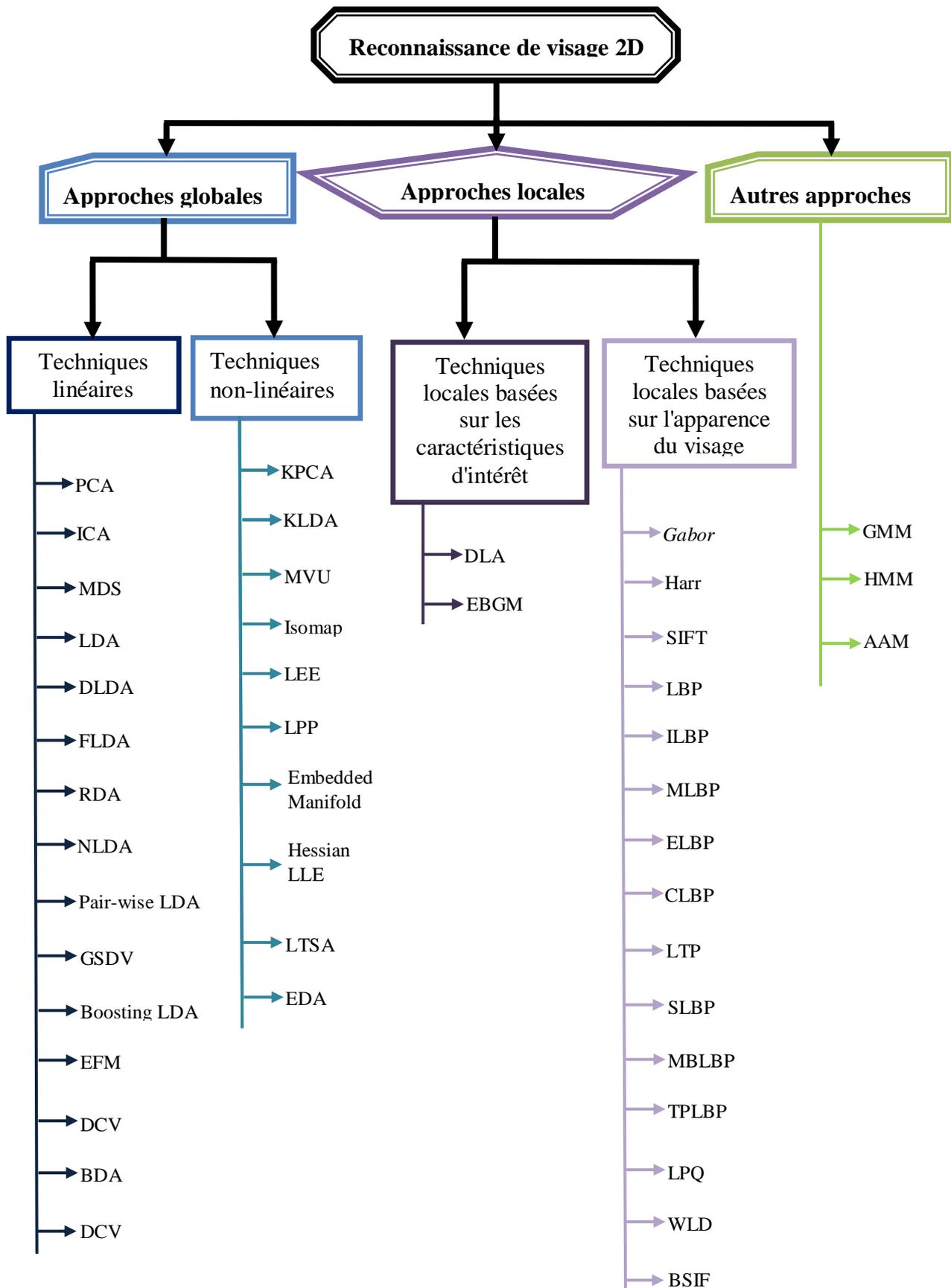


Figure 2.12: Classification des principaux algorithmes utilisés en reconnaissance faciale 2D.

2.2.4 Discussion

Nous avons présenté dans cette section l'état de l'art en reconnaissance de visage 2D. Plusieurs algorithmes appartenant à deux catégories principales: les approches globales et locales dans la reconnaissance du visage (les autres approches étant basées sur les modèles statistiques). En comparant ces approches, les méthodes globales ont tendance à être moins sensibles au bruit. Cependant, les méthodes locales ont montré leur performance prometteuse au cours des dernières années. Il a été prouvé par *Heisele et al.* [181] que la reconnaissance de visage à base des méthodes locales est plus performante que celle des méthodes globales. La raison principale est que les approches holistiques exigent toujours des images du visage pour être normalisées avec précision en ce qui concerne les facteurs de pose, l'état de l'éclairage et de l'échelle. En outre, les caractéristiques globales sont également plus sensibles aux variations d'expression et des occlusions faciales. Selon sa nature théorique, AAM peut également être considérée comme un exemple particulier de techniques holistiques et elle possède de leurs avantages et inconvénients dans la reconnaissance des visages. Parce que les techniques basées sur les caractéristiques locales permettent d'extraire des points caractéristiques locaux, il reste généralement quelques caractéristiques invariantes, même en présence de variations d'expression du visage ou des occlusions et la reconnaissance peut encore être faite en faisant correspondre ces caractéristiques invariantes. Par conséquent, les méthodes basées sur les caractéristiques locales sont potentiellement plus robustes que celles holistiques à des changements d'expression et des occlusions faciales. En outre, à la différence des approches holistiques, elles nécessitent moins d'échantillons d'apprentissage et peuvent même atteindre une excellente précision avec une seule image du visage enregistrée dans l'ensemble de la galerie [182]. Il existe également plusieurs tâches, combinant des techniques globales avec des techniques locales et visant à fusionner leurs avantages, mais en même temps elles héritent des inconvénients des deux types de catégories.

On pourrait poser la question: qui est le meilleur descripteur local pour la reconnaissance faciale chez LBP, SIFT et filtres de *Gabor*? *Luo et al.* [183] ont montré que SIFT n'est pas aussi robuste que LBP à l'éclairage des effets sur l'ensemble de données FERET. *Zou et al.* [70] ont comparé les ondelettes de *Gabor* et LBP utilisant la même base de données et a conclu que les ondelettes de *Gabor* sont plus sensibles aux variations d'éclairage, car elles détectent des fréquences spatiales invariantes de valeurs de gris de pixels de l'image, tandis que LBP est grandement affectée par les transformations de valeurs de gris non-monotones. *Ruiz-del-Solar et al.* [184] ont largement évalué ces trois méthodes, non seulement sur des bases de données comme FERET et FRGC, mais aussi sur la base de données UCH Face HRI conçue pour l'interaction homme-robot et l'ensemble de données LFW capturés dans des environnements contraignants. En ce qui concerne la robustesse aux variations d'éclairage, leur étude a montré que les ondelettes de *Gabor* atteignent la meilleure performance sur la base de données FERET. La LBP n'est pas loin derrière, tandis que la méthode SIFT est la moins performante.

Sur la base de données UCH Face HRI, l'approche LBP donne les meilleurs résultats dans toutes les expériences spécialement conçues avec éclairages intérieur et extérieur, expression, mise à l'échelle et rotation. Elle est suivie par les ondelettes de *Gabor* et SIFT. Sur l'ensemble de données LFW, la LBP et les ondelettes de *Gabor* obtiennent un résultat légèrement meilleur avec le visage aligné et non-aligné. Tous ces deux descripteurs dépassent la SIFT. D'autre part, dans leur enquête sur le coût de calcul, la LBP fonctionne beaucoup plus vite que les ondelettes de *Gabor* et SIFT. Il est difficile de tirer une conclusion générale pour classer ces descripteurs locaux, car leur rendement est déterminé par un certain nombre de facteurs complexes, lequel devrait être sélectionné pour la reconnaissance des visages qui dépend des exigences des applications.

2.3 Reconnaissance de visage 3D

Avec le développement des scanners 3D et les techniques de capture, de nombreuses approches 3D sont apparues. Semblables à la taxonomie de reconnaissance faciale 2D, les approches de

reconnaissance de visage 3D peuvent également être classées en deux approches: globales et locales. Dans cette section, les travaux antérieurs qui utilisent uniquement la forme 3D pour la reconnaissance faciale sont examinés. Beaucoup de recherches antérieures avaient accès à de très petites bases de données et, par conséquent, elles signalent des taux de reconnaissance étonnamment élevés. Au fil du temps, les enjeux sont devenus plus grands et les algorithmes ont produit des taux plus précis.

2.3.1 Approches globales

Comme pour les approches globales basées sur les images 2D qui exploitent la totalité de l'image 2D du visage en entrée du système, les informations géométriques 3D correspondant aux approches globales travaillent directement sur toute l'image de profondeur ou nuage de points du visage pour calculer la mesure de similarité.

2.3.1.1 Le sous-espace (PCA et LDA)

Les méthodes basées sur le sous-espace sont généralement appliquées à des images de profondeur pour le visage. *Achermann et al.* [185] ont appliqué l'approche du visage propre à la reconnaissance faciale 3D. Ils présentent les résultats d'un ensemble de données de 24 sujets qui ont chacun 10 images et trouvent un taux de reconnaissance de 100%.

Plus tard, *Hesher et al.* [186] explorèrent la PCA en utilisant des nombres différents de vecteurs propres. Les données utilisées sont 6 expressions faciales différentes pour chacune des 37 personnes. Les résultats de l'utilisation de plusieurs images du visage par personne dans l'ensemble de la galerie donnent effectivement à l'image de test plus de chances de faire une similarité correcte.

Heseltine et al. [187] évaluèrent l'approche de visage propre avec différentes distances comme la mesure de similarité pour la vérification du visage sur une grande base de données contenant 330 images de plus de 100 personnes. Ils conclurent que le plus faible taux d'erreur égalité (TEE) 17,8% était atteint pour la distance de Mahalanobis et non avec les distances euclidienne et cosinus.

Heseltine et al. [188] élargirent l'ensemble de test à 1470 images de profondeur de 230 personnes et exploitèrent la LDA dans le même cadre. Ils affirmèrent que la LDA est plus performante que la PCA avec la distance cosinus. Avec la LDA, ils obtinrent un TEE= 15,3%.

2.3.1.2 Iterative Closest Point (ICP)

L'algorithme ICP [189] tente de manière itérative d'aligner deux surfaces 3D représentées comme point de nuages ou mailles. Pour accomplir cette tâche, l'ICP trouve d'abord le point de la surface de référence le plus proche pour chacun des n points de test.

ICP est largement utilisée pour l'enregistrement de visage et la correspondance à la reconnaissance faciale 3D. *Medioni et Waupotitsch* [190] ont effectué une reconnaissance de visage 3D en utilisant ICP pour la correspondance des surfaces faciales. Alors que la plupart des travaux utilisent des formes 3D acquises par un capteur à lumière structurée (structured-light), ce travail utilise des formes 3D acquises par un capteur stéréo passif (passive stereo). La base de données contient 100 sujets, dont chacun possède 7 images de pose différentes. Un TEE mieux que 2% est signalé.

Wang et al. [64] ont proposé l'ICP partielle qui sélectionne une partie de paires de points de proximité pour calculer la mesure de similarité lors de l'enregistrement des surfaces du visage et afin de réduire les effets négatifs causés par les changements d'expression du visage. Les expériences sur un ensemble de données de 360 modèles de visage à partir de 40 personnes montrent que l'ICP partielle est meilleure que l'ICP originale.

Dans la plupart des travaux récents, l'ICP est largement présentée comme une étape clé dans la phase de prétraitement pour corriger les variations de pose 3D comme dans [191, 192].

Mohammadzade et Hatzinakos [193] ont proposé la méthode "Iterative Closest Normal Point" pour trouver les points correspondants entre une face de référence générique et chaque face d'entrée. Les

propositions de la méthode de correspondance sont la recherche des échantillons d'un ensemble de points pour chaque visage. Il faut noter que les points les plus proches sont normaux. Ces points sont effectivement alignés sur tous les visages, permettant une application efficace des méthodes d'analyse discriminante pour la reconnaissance faciale 3D. Par conséquent, le problème de la variation de l'expression est traité en minimisant la variabilité intra-classe des échantillons pour le visage, tout en maximisant la variabilité inter classe. Comme conclusion, les auteurs ont montré que les vecteurs normaux de surface du visage aux points échantillonnés contiennent des informations plus discriminatoires que les coordonnées des points. Les performances globales des expériences sur la base de données FRGC v2 atteignent des taux de vérification de 99,6 % et 99,2 % et un taux de fausses acceptations de 0,1 % pour "le tout contre tous" et ROC III, respectivement.

Les figures 2.13, 2.14 et 2.15 illustrent avec trois exemples différents les résultats de l'application de l'algorithme *ICP* pour le recalage des surfaces faciales [194].

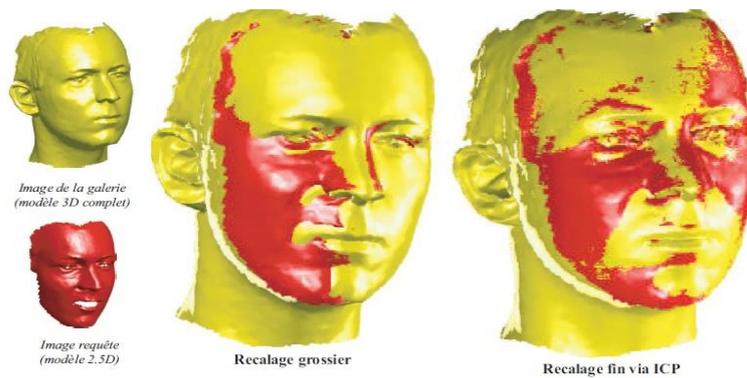


Figure 2.13: Exemple 1: recalage entre un modèle 3D et un modèle requête de la même personne avec variations d'expressions faciales [194].

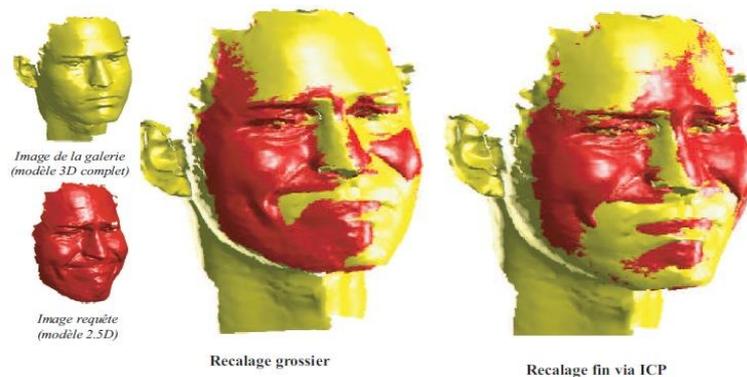


Figure 2.14: Exemple 2: recalage entre un modèle 3D et un modèle requête de la même personne avec des variations d'expressions faciales plus significatives [194].

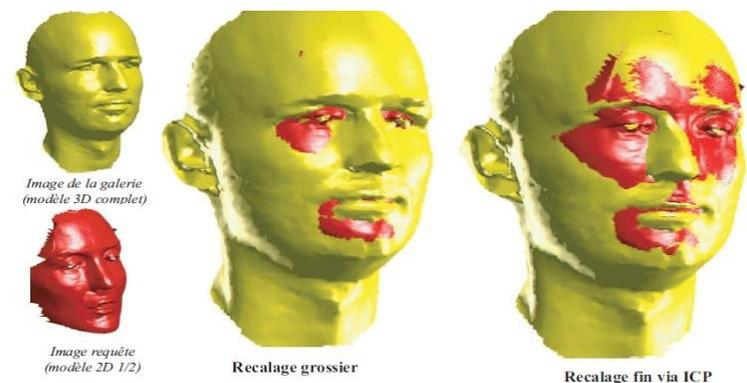


Figure 2.15: Exemple 3: recalage entre un modèle 3D et un modèle requête de deux personnes différentes [194].

2.3.1.3 Distance de Hausdorff

La distance de Hausdorff est une mesure de la similarité de deux ensembles de points [195]. Elle est très générale et peut être appliquée à une grande variété de problèmes. La distance non dirigée de Hausdorff entre deux ensembles de points A et B est définie comme:

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (2.18)$$

Où $h(A, B)$ désigne la distance de Hausdorff dirigée,

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (2.19)$$

$\|x\|$ est la fonction norme. Habituellement, la distance euclidienne bien connue est utilisée ici. La distance de Hausdorff partielle est définie comme:

$$H_{LK}(A, B) = \max(h_L(A, B), h_K(B, A)) \quad (2.20)$$

Où la distance de Hausdorff partielle dirigée est définie comme :

$$h_L(A, B) = L_{a \in A}^{th} \min_{b \in B} \|a - b\| \quad (2.21)$$

Nous remarquons que $h(A, B)$ est petite, si chaque point A est à proximité de certains points B . Il en est de même pour $h(B, A)$. En outre, $H(A, B) = H(B, A)$ est faible, si à la fois $h(A, B)$ et $h(B, A)$ sont petites. Dans la distance de Hausdorff partielle, seule les points L et K mieux classés en séries A et B sont pris en compte. Ainsi, il est possible d'ignorer les valeurs aberrantes et gérer les cas où une seule partie des données correspond à un modèle donné, ou une partie d'un modèle correspond à un ensemble donné de points de données.

La distance de Hausdorff a été utilisée à l'origine pour des ensembles de points dans le plan 2D [195]. Cependant, les équations 2.18 et 2.21 peuvent être appliquées à des points dans l'espace 3D et il est nécessaire de faire l'apprentissage à l'avance des deux au moins pour la comparaison, c'est à dire, de trouver une translation et une rotation qui alignent de manière optimale les deux ensembles de données.

Achermann et *Bunke* [196] utilisent une extension de la distance de Hausdorff pour la reconnaissance de visage 3D. Ils rendent compte des expériences utilisant 240 images de profondeur, 10 images de chacune des 24 personnes et atteignent la reconnaissance de 100% pour certains cas.

Pan et al. [197] ont comparé une approche de la distance de Hausdorff et celle fondée sur la PCA en reconnaissance faciale 3D. Dans les expériences avec des images de la base de données 3D_RMA, ils signalent un TEE de l'ordre de 3% à 5% pour l'approche de la distance de Hausdorff et 5% à 7% pour l'approche fondée sur la PCA.

Lee et Shim [198] proposent une «depth-weighted Hausdorff distance" pour calculer le score de similarité qui est ensuite fusionné avec l'information de courbure de la surface (le minimum, le maximum et la courbe de Gauss) pour la reconnaissance faciale 3D. Ils présentent les résultats des expériences avec un ensemble de données représentant 42 personnes, avec deux images chacune. Le taux de reconnaissance est de 98% pour la meilleure méthode de combinaison, tandis que la plaine distance de Hausdorff réalise moins de 90%.

Russ et al. [199] ont exploré la correspondance sur les images 3D du visage. Dans une expérience de vérification, 200 personnes sont inscrites dans la galerie et les mêmes 200 personnes et 68 autres imposteurs pour le test. Une probabilité de vérification correcte de 98% (pour 200 personnes) est réalisée à un taux de fausse alarme de 0 (pour les 68 imposteurs).

2.3.1.4 Modèle déformable

Les visages humains ne sont pas rigides et leurs variations d'expression peuvent beaucoup déformer leurs surfaces, ce qui rend très difficile l'utilisation des méthodes globales mentionnées ci-dessus. Le modèle déformable est proposé afin de mieux décrire les transformations non-rigides des visages. En utilisant des échantillons représentatifs dans la base d'apprentissage, le modèle déformable peut être utilisé pour [6]:

- l'apprentissage de la quantité de déformations entre les différentes expressions du visage en indiquant si elles appartiennent à une classe intra variations,
- l'apprentissage des transformations entre différentes expressions faciales pour agrandir l'ensemble de test avec des modèles de visage plus expressifs ou pour transformer un visage expressif de la base de test à un modèle neutre pour la reconnaissance.

Lu et Jain [200] étendent leur travail précédent en utilisant une approche de reconnaissance à base de ICP [201] pour traiter explicitement des variations d'expression du visage. Le problème est résolu comme une transformation rigide du modèle de test de la galerie avec l'ICP en utilisant une déformation non-rigide conçue par des techniques TPS (Thin-Plate Spline). L'approche proposée est évaluée en utilisant un ensemble de données de 100 personnes, avec expression neutre dans l'ensemble galerie et expression sourire dans l'ensemble de test. Les entrées de la galerie sont des structures de données frontales, alors que toutes les données de test sont des vues frontales. La plupart des erreurs sont enregistrées après résultat de la transformation rigide du sourire et ces erreurs sont réduites après l'étape de déformation non rigide.

Lu et Jain [202, 203] proposent une autre méthode de reconnaissance faciale 3D qui est robuste par rapport aux variations d'expression du visage. Basée sur l'information apprise en utilisant le groupe de contrôle pour chaque sujet, l'approche peut générer différents modèles de visage d'expression neutre dans l'ensemble de la galerie. La figure 2.16 donne une illustration de l'ensemble de ce travail. Un petit groupe de contrôle constitué de 10 sujets, chacun avec sept expressions (neutre, heureux, en colère, sourire, la surprise, dégonflé et gonflé), est recueilli à partir de la base de données MSU. Bien que l'ensemble de test contienne 877 images de 100 personnes choisies de la base de données FRGC v2, les résultats expérimentaux montrent que l'approche proposée est capable de gérer les changements d'expressions dans une certaine mesure.

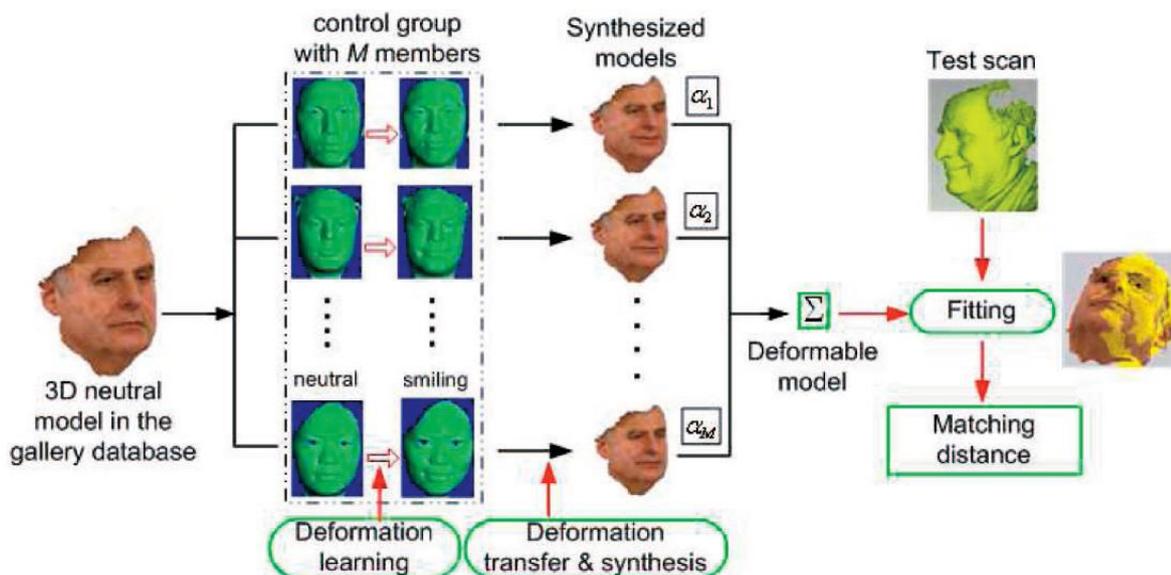


Figure 2.16: Synthèse et transfert de la déformation basés sur la reconnaissance faciale 3D [203].

Al-Osaimi et al. [204] ont pris les modes de déformations d'expression à partir des données de formation en vecteurs propres de PCA et ces modes sont ensuite utilisés pour transformer les déformations d'expression. La PCA est réalisée de telle manière qu'elle ne modélise que les expressions faciales en laissant de côté les disparités interpersonnelles. Des mesures de similarité sont extraites en faisant correspondre les visages 3D transformés. L'approche est appliquée sur la base de données FRGC v2. Elle a atteint le taux de vérification de 98,35% et 97,73% à 0,001 TFR pour les analyses dans le cadre des expressions neutres et non neutres, respectivement. La figure 2.17 présente certains vecteurs propres qui couvrent la forme sous-espace PCA (Expression Deformation Model, EDM). Les images de la première rangée sont les vecteurs propres 1, 2 et 3, respectivement à partir de la gauche. Les vecteurs propres 9, 10 et 11 sont dans la deuxième rangée. Les vecteurs propres 25, 26 et 27 sont dans la troisième rangée du bas. Notons que les meilleurs vecteurs propres sont plus pertinents pour les régions largement déformables (par exemple la bouche) et les déformations dans les autres régions apparaissent progressivement dans les vecteurs propres inférieurs.

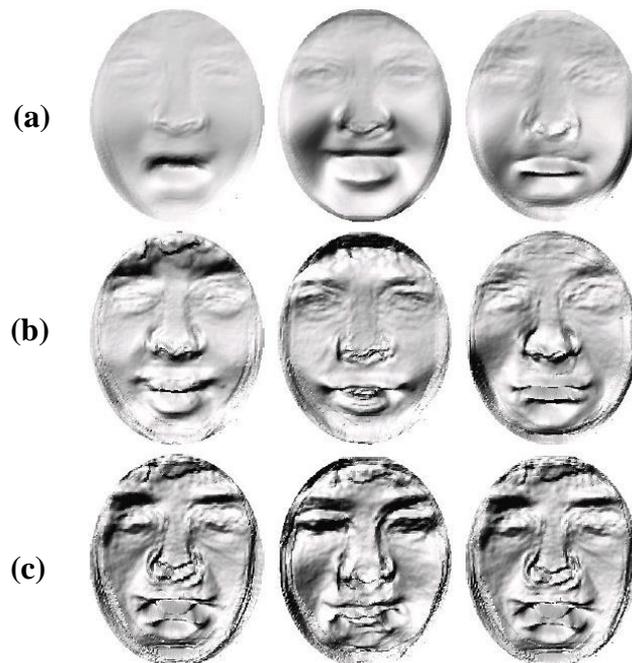


Figure 2.17: (a) vecteurs propres 1, 2 et 3, respectivement à partir de la gauche, (b) vecteurs propres 9, 10 et 11 et (c) vecteurs propres 25, 26 et 27 [204].

2.3.2 Approches locales

Les approches locales ont été largement étudiées dans la reconnaissance faciale 3D car elles procurent généralement de meilleurs résultats que les approches globales. Ces approches sont classées dans [81] en cinq classes.

2.3.2.1 Méthodes basées sur le calcul du point

Les méthodes basées sur le calcul du point extraient des informations des caractéristiques d'un ensemble de points initiaux pour la correspondance, comme: les coordonnées originales, la distance d'un bord entre deux points, la surface d'une région composée de plusieurs points, l'angle entre les deux bords, exigent des descripteurs plus complexes.

Wu et al. [205] proposent LSM (Local Shape Map) pour décrire un point sur la surface du visage. Le coefficient de corrélation est utilisé en tant que mesure de similarité pour comparer les LSM tous les deux points et les résultats sont ensuite incorporés dans une méthode de vote pour la classification.

Cette méthode ne nécessite pas d'enregistrer deux surfaces faciales. Avec une base de données de 31 images de profondeur à partir de 6 sujets, Wu et al. trouvent un TEE de 2,98%.

Castellani et al. [206] introduisent une méthode d'apprentissage générative par adaptation de HMM qui travaille sur des maillages 3D. La géométrie de la zone locale autour des points de référence est modélisée par la formation de HMM, qui fournit un point robuste invariant lors de la rotation. Cette description permet de comparer la signature des points correspondants par maximum de vraisemblance. Selon la performance préliminaire sur un petit sous-ensemble de FRGC v2, les auteurs affirment que l'approche proposée est robuste face aux variations d'expression du visage et fonctionne plus rapidement que les systèmes de reconnaissance faciale 3D actuels basés sur l'ICP qui maintient un taux de reconnaissance satisfaisant.

Daniyal et al. [207] forment un modèle dit du point de distribution (Point Distribution Model, PDM) en localisant manuellement des points sur 100 exemples de modèles et en localisant automatiquement sur l'ensemble de test. Toutes les distances euclidiennes entre tous les deux points sont exploitées en tant que caractéristiques et classées par la méthode LDA pour la reconnaissance 3D. L'approche proposée est évaluée sur la base de données BU-3DFE et atteint un taux de reconnaissance de 96,5%.

Maes et al. [208] étendent l'algorithme SIFT à ce qui est connu sous le nom Mesh SIFT. Contrairement aux informations du gradient extraites par l'opérateur SIFT original, Mesh SIFT code l'histogramme local de l'indice de forme. Les tests sont effectués sur l'ensemble des données de Bosphorus et atteignent un taux de reconnaissance égal à 93,7%.

Berretti et al. [209] ont proposé un système de reconnaissance de visage 3D dans le cas où seulement des parties des scans de test sont disponibles. La figure 2.18 présente des points-clés détectés sur les scans du visage de cinq personnes différentes pour la combinaison de la base de données FRGC v2 et UND. Chaque colonne représente trois scans du visage de la même personne: le scan de la galerie (FRGC v2 gallery), un balayage de test frontal (FRGCv2 probe) et un balayage de test de côté (UND left/right probe). On peut remarquer qu'une grande partie des points-clés sont identifiés de façon répétée dans les mêmes quartiers pour la même personne.

2.3.2.2 Méthodes basées sur le calcul de courbe

Les méthodes basées sur le calcul de courbe comparent les courbes correspondantes de différentes surfaces faciales 3D ou calculent les mesures de similarité en utilisant des fonctionnalités plus robustes.

Samir et al. [210] ont présenté une nouvelle approche de comparaison de surfaces faciales 3D pour la reconnaissance. Dans un premier temps, les auteurs donnent une nouvelle représentation faciale basée sur les courbes de niveaux (level curves) extraites à partir des images de profondeur. Un certain nombre de prétraitements était introduit avant cette opération, tels que la conversion du modèle 3D en une image de profondeur, le bouchage des trous, l'élimination des pics par lissage gaussien et finalement la segmentation par un masque elliptique de la surface faciale. Ils présentèrent une métrique de comparaison entre ces représentations, appelées aussi descripteurs. Cette métrique est basée sur une approche de calcul de chemin géodésique qui est décrite dans [211] entre les points appartenant aux mêmes courbes de niveaux qui ont été extraites dans la première étape.

Zhang et al. [212] ont fait usage de modèles de visage 3D et leurs images pour extraire les profils de symétrie centrale. Sur la base des valeurs de courbure des profils, trois points essentiels sont situés sur chaque surface du visage. Ces trois points déterminent le FICS (Face Intrinsic Coordinate System) qui est utilisé pour aligner tous les visages. Le profil de symétrie ainsi que deux profils transversaux composent une représentation compacte pour le visage correspondant. La méthode proposée est testée sur 382 surfaces de face, venant de 166 personnes. Les valeurs de TEE pour l'authentification des visages neutres et expressifs sont de 0,8 % et 10,8 %, respectivement. L'article publié dans PAMI (Pattern Analysis and Machine Intelligence) [211] contient une bonne évaluation de l'approche proposée. Cette évaluation est effectuée sur deux bases de données: la base FSU collectée par le biais

d'un scanner Minolta VI- 700 dans un contexte contrôlé et une partie de la base UND (University of Notre Dame) dont l'acquisition est effectuée par le scanner VI-900/910 de Minolta. Sur la première base de visages (FSU data base), un meilleur taux de reconnaissance égal à 92% est obtenu, en expérimentant avec un ensemble de modèles représentant différentes expressions faciales. Les auteurs concluent également à une augmentation du taux de reconnaissance en fonction du nombre de courbes de niveaux utilisées. Sur la même base de test, les auteurs comparent leur approche à une approche de type visages propres appliquée sur les images de profondeur. A l'issue de cette étude comparative, ils affirment réaliser des taux de reconnaissance plus intéressants que ceux de la méthode de référence. Concernant la deuxième base d'expérimentation, uniquement un sous-ensemble de 740 scans (162 sujets) est gardé de la base initiale UND. Il est divisé en 470 images de la galerie et 270 images de test (probe images). Un taux de 90,4 % de reconnaissance est obtenu pour un nombre optimal de courbes de niveaux égal à 6. Au-delà de ce nombre, aucune amélioration n'est apportée. Les taux de reconnaissance obtenus valident l'aspect théorique approfondi de cette approche. Néanmoins, elle est réservée aux images frontales du visage comme le montrent les expérimentations.

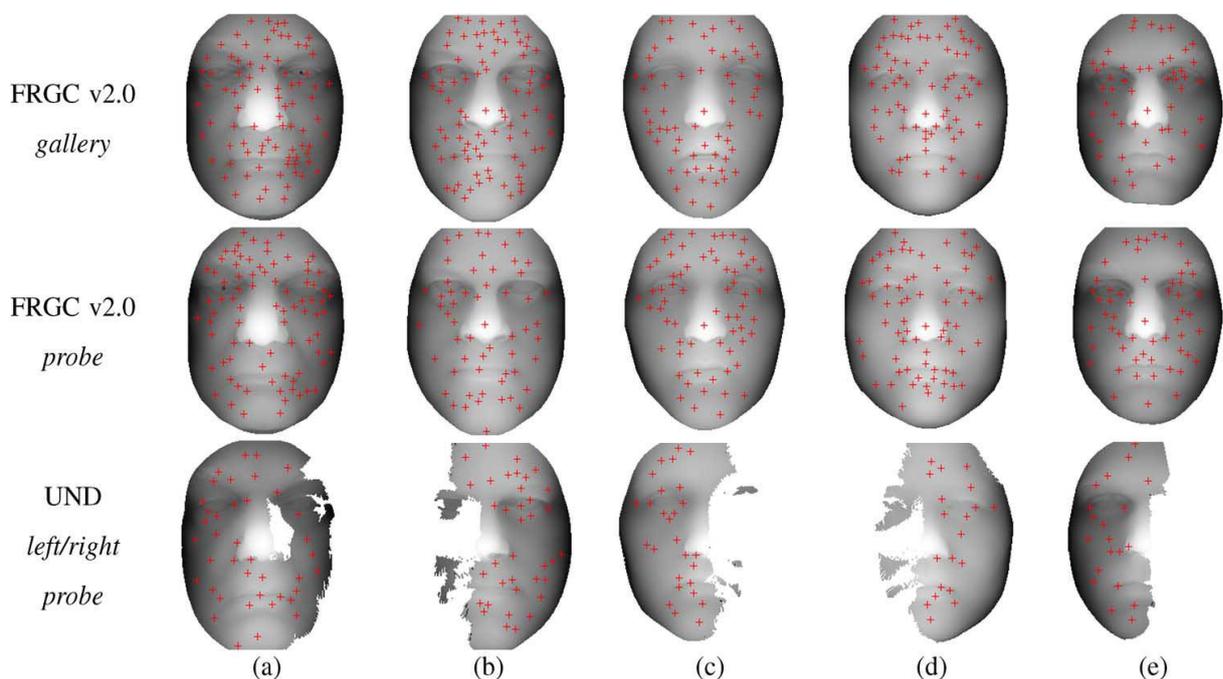


Figure.2.18: Points-clés détectés sur les scans du visage de cinq personnes différentes pour la combinaison de la base de données FRGC v2 et UND [209].

Mahoor et al. [68] introduisent la courbure principale, k_{max} , pour représenter l'image du visage comme une image 3D binaire appelée image de crête. L'image de crête indique l'emplacement des lignes de crête autour des régions faciales importantes sur le visage (les yeux, le nez et la bouche). Les expériences sur l'ensemble de données Gavab composé de 61 sujets ont donné un taux de reconnaissance de 93,5 % pour une expression neutre et 82,0% pour les visages avec une expression de sourire.

Jahanbin et al. [213] décrivent les surfaces 3D par iso-profondeur et courbes iso-géodésique. La première est produite par l'intersection d'une surface faciale avec des plans parallèles perpendiculaires à la direction du regard, à des profondeurs différentes de la pointe du nez. Celui-ci est défini comme étant le lieu géométrique de tous les points de la surface du visage avec la même distance à partir d'un repère géodésique facial donné. Une fois que les courbes du visage sont extraites, leurs caractéristiques sont codées par plusieurs fonctionnalités telles que descripteurs de forme simple (convexité, rapport des axes principaux, compacité, variance et variance circulaire ou elliptique) ou les distances euclidiennes polaires à partir de l'origine. La dernière étape consiste à vérifier ou rejeter les demandes des utilisateurs à l'aide d'un classificateur de distance euclidienne ou SVM. Les expériences sur un

groupe de données contenant 1196 modèles de 119 sujets montrent que le rendement de l'iso-géodésique est bien meilleur que celui de l'iso-profondeur.

Drira et al. [214] ont exploré les courbes radiales élastiques (Elastic Radial Curves) pour modéliser les déformations faciales 3D causées par les variations d'expression du visage. Les distances géodésiques entre les courbes sont adoptées comme mesures de similarité pour comparer les surfaces faciales. Cette technique permet d'obtenir un taux de reconnaissance global de plus de 90% sur l'ensemble des données Gavab. Figure 2.19 illustre le processus de déformation correspondant et la distance géodésique de courbes radiales.

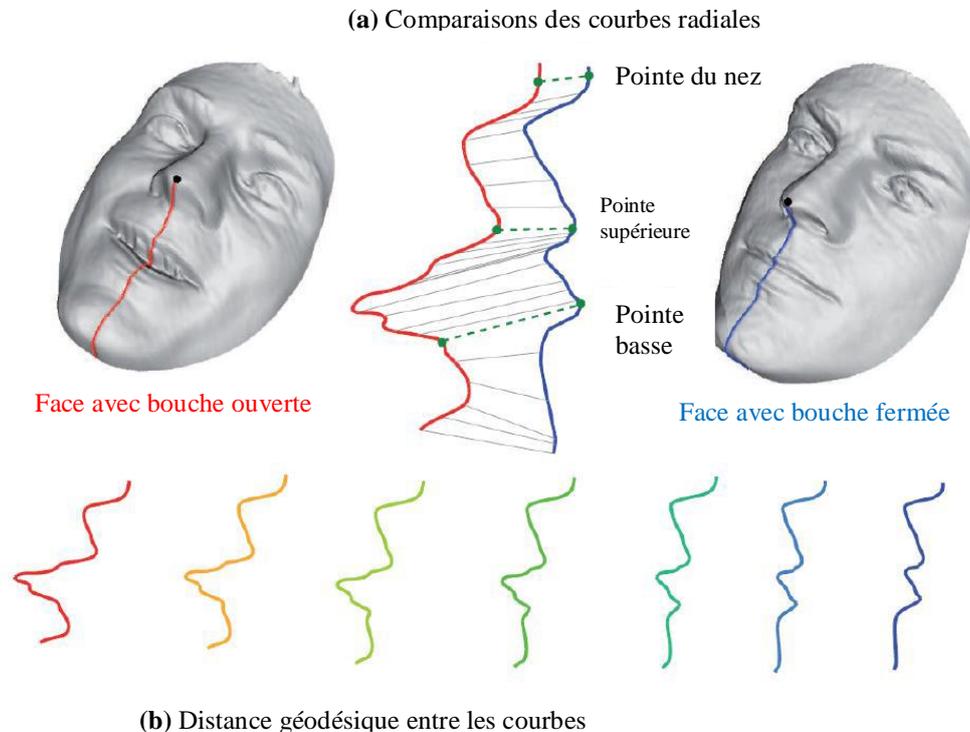


Figure 2.19: Comparaisons des courbes radiales et la distance géodésique entre les courbes [213].

2.3.2.3 Méthodes basées sur les régions

La totalité de la surface du visage peut être segmentée relativement en parties rigides et non rigides en attribuant un poids plus faible sur les régions du visage non rigides pour réduire leurs effets. Les régions du visage sélectionnées sont ensuite comparées à un certain espace de caractéristiques pour la reconnaissance faciale.

Alyuz et al. [215] proposent une approche régionale de l'enregistrement qui est robuste face aux expressions pour la reconnaissance faciale 3D. Comme le montre la figure 2.20, cette méthode se compose de quatre parties:

- la localisation du repère automatique,
- l'enregistrement du visage global et local basé sur l'ICP,
- l'extraction de caractéristiques du visage,
- la fusion par classification des régions.

Dans la première partie, cinq points de référence sont automatiquement localisés en utilisant des méthodes basées sur la courbure. Ces points sont ensuite utilisés pour l'alignement de la face globale basé sur la méthode dite "Procrustes Analysis". La méthode d'enregistrement de la région sur la base proposée est inspirée par le modèle du visage moyen (AvFM) [215, 216] et étendue aux régions locales indépendantes, ce qui est appelé le modèle de la région moyenne (AvRM). Le AvRM utilise 15 correctifs locaux, dont 7 régions significatives. Une autre région qui couvre des correctifs locaux

appartenant à l'œil, le nez et les zones de front, en termes face supérieure, est également construite. L'inscription en AvRM fournit la correspondance de tous les points de la surface du visage un-à-un. Deux types de descripteurs de forme 3D: coordonnées des points et courbures de surface faciale sont considérés. L'algorithme du plus proche voisin est utilisé. Deux bases de données du visage 3D: la FRGCv2 et le sous-ensemble de base de données Bosphorus hors variation de pose et d'occlusion sont utilisés pour l'évaluation expérimentale. Pour la base de données FRGCv2, le meilleur classement avec un taux de reconnaissance égale 94,80 % est obtenu par fusion de 16 classificateurs individuels (8 pour chaque type de descripteur) avec le système de scrutin majoritaire modifié. Pour la base de données du Bosphorus, un taux de reconnaissance de 98,08 % est obtenu avec la fusion de deux types de descripteurs et 8 régions locales. Pour améliorer encore le rendement, l'analyse statistique (LDA) est en outre effectuée sur la base de la fonction pour la réduction de la dimensionnalité.

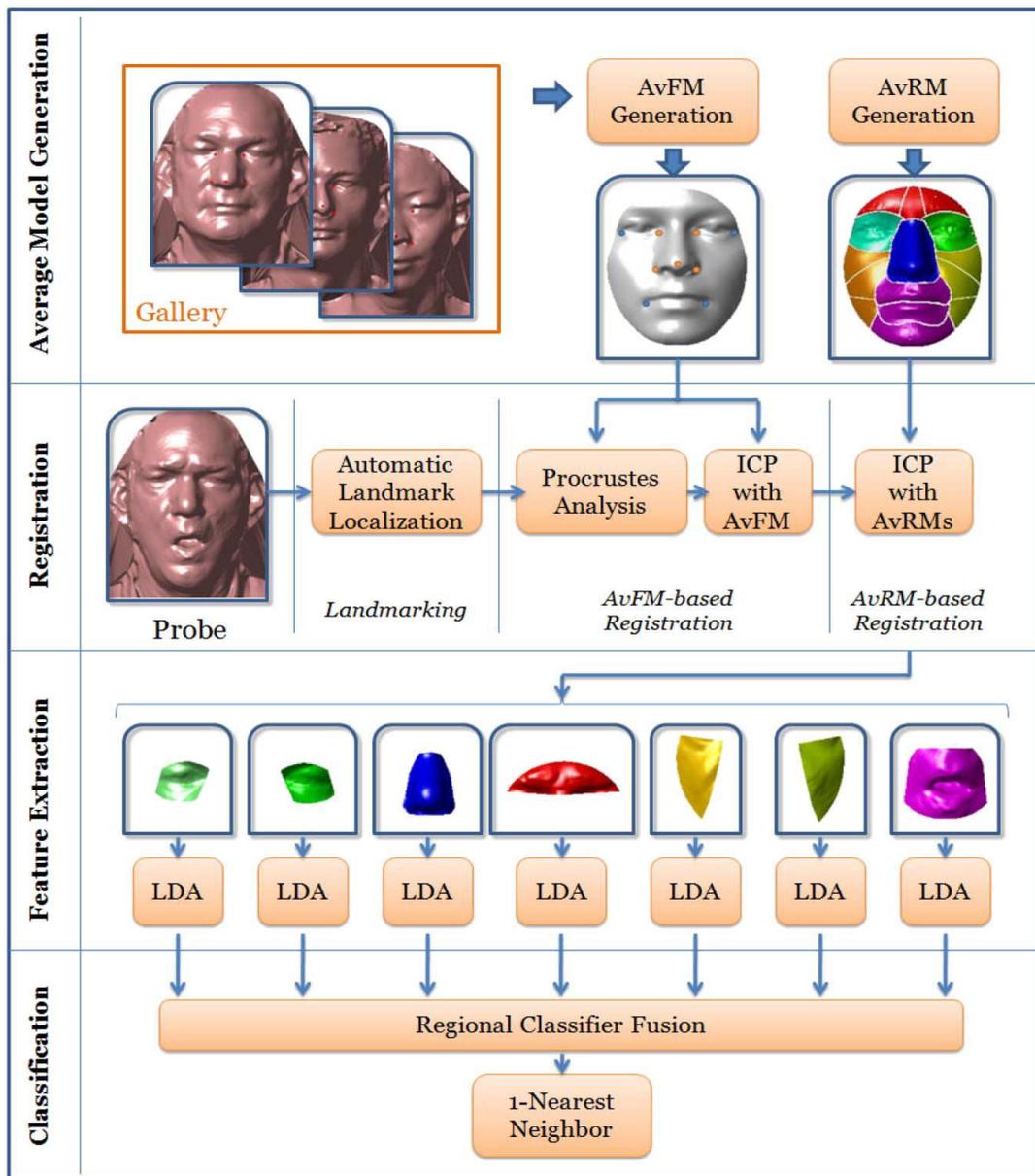


Figure 2.20: Vue d'ensemble de reconnaissance 3D de visage approchée fondée sur l'enregistrement régional [218].

Différentes méthodes basées sur l'enregistrement du modèle du visage, *Spreeuwers* [218] proposent un procédé de reconnaissance faciale 3D rapide et précis qui effectue l'enregistrement du visage par rapport à un système de coordonnées intrinsèques. L'origine du système de coordonnées intrinsèques est définie par la pointe du nez et les trois angles du système de coordonnées sont déterminés par le plan de symétrie et de l'angle de l'arête du nez. L'algorithme RANSAC (Random Sample Consensus) [219] est utilisé pour effectuer le raccord cylindre et la ligne du corps pour la détection de visage et l'estimation de la pente de pont du nez. Le nuage de points alignés est utilisé pour générer l'image de profondeur à haute résolution par ré-échantillonnage. Pour faire face au problème de l'expression, les auteurs ont défini un ensemble de 30 régions de chevauchement qui sont présentées dans la figure 2.21 où la zone blanche est incluse et la zone noire est exclue. Pour comparer deux images, plusieurs classificateurs de la région sont fusionnés pour calculer la similarité. La forme de vectorisation de chaque région est utilisée pour construire le vecteur de caractéristiques, dont la dimension est encore réduite en effectuant une PCA suivie par une LDA. Ensuite, un rapport de vraisemblance est utilisé pour la classification régionale. Enfin, la stratégie de fusion à base de vote à la majorité est utilisée pour la fusion. Pour l'évaluation expérimentale, la base de données de Bosphorus (toutes les analyses frontales sont sans occlusions) et la base de données 3DFACE sont utilisées pour la phase de formation des classificateurs de la région, tandis que la base de données FRGCv2 est utilisée pour la phase de test.

La figure 2.22 montre les scores de rang un pour 30 classificateurs de la région formée sur la base de données frontale Bosphorus. Le maximum de scores de rang un obtenus pour un seul classificateur est de 95,9 %, tandis qu'un score de rang un de 97,9 % est rapporté par la fusion de tous les 30 classificateurs. Des résultats similaires peuvent être obtenus pour les classificateurs de la région formée sur la base de données 3DFACE. Enfin, l'auteur fait état d'un score de rang un de 99,0 % par fusion de deux fois 30 plus. C'est la plus haute performance obtenue. Le coût du temps est aussi une supériorité de cette méthode. Il ne faut que 2,5 secondes pour identifier une seule image à l'aide d'une galerie de 466 sujets de la base de données FRGCv2. Cette méthode est donc plus de 700 fois plus rapide que la méthode de [220].

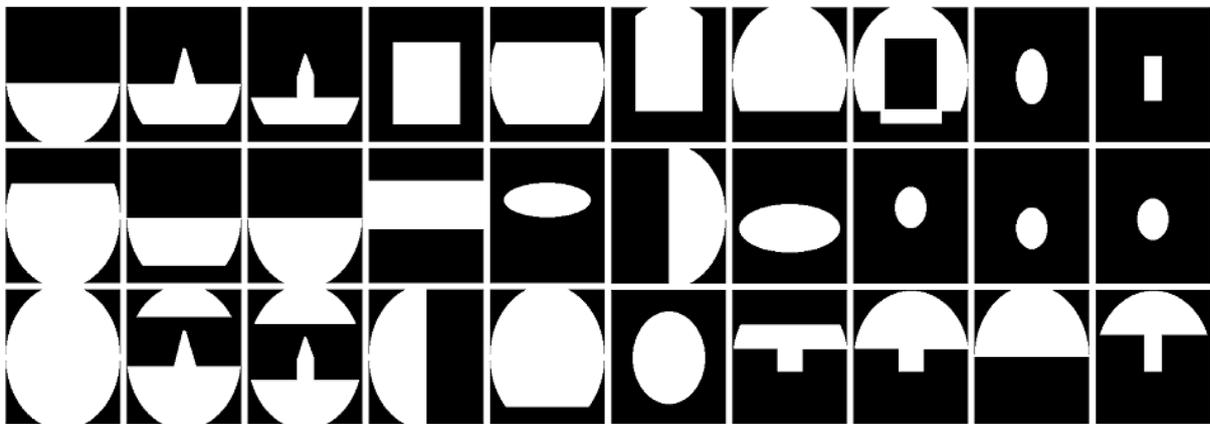


Figure 2.21: 30 régions locales utilisées pour la comparaison du visage [219].

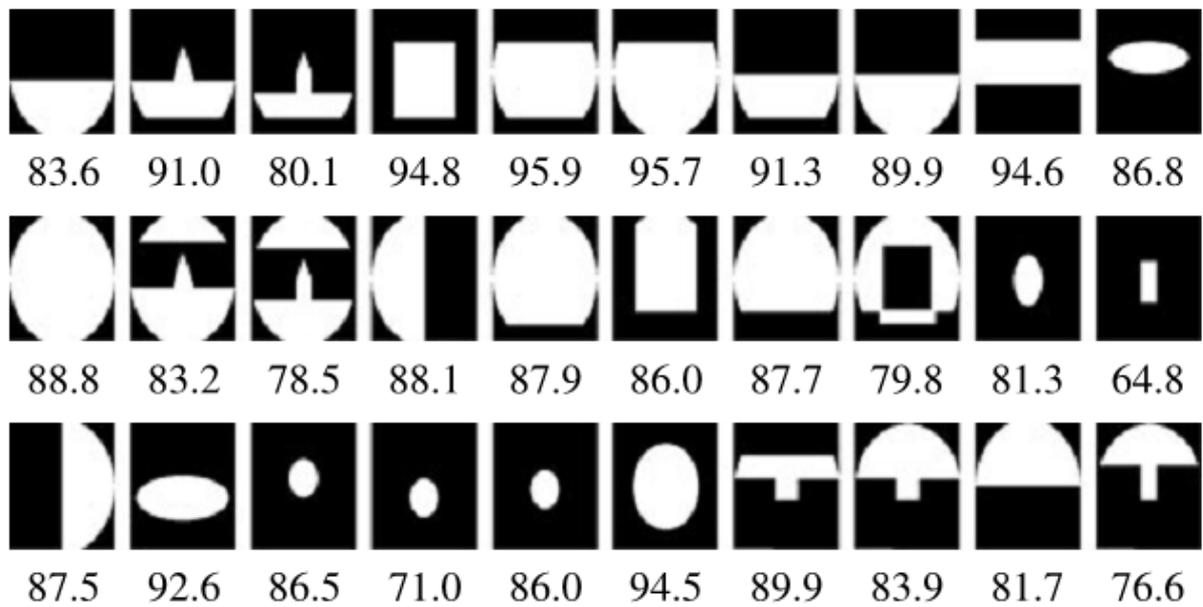


Figure 2.22:Taux de reconnaissance des régions individuelles sur la base de données FRGC v2 [219].

2.3.2.4 Les méthodes basées sur les descripteurs

Pour représenter chaque sommet d'un nuage de points du visage 3D (mesh) ou chaque pixel d'une image de profondeur pour le visage, les descripteurs généraux sont des méthodes pour extraire localement des attributs de forme de son voisinage. L'ensemble du visage peut être décrit par un vecteur caractéristique combinant les informations de ces descripteurs qui est enfin introduit dans la phase de classification.

Gupta et al. [221] proposent d'utiliser les ondelettes complexes à similarité structurelle métrique (Complex Wavelet Structural Similarity Metric, CW-SSIM) extraites des images de profondeur pour le visage pour la reconnaissance faciale 3D. La CWSSIM calcule efficacement et d'une façon robuste aux petites distorsions géométriques. Comparée avec l'erreur quadratique moyenne (Mean Squared Error, MSE) et la distance de Hausdorff, CW-SSIM offre de meilleures performances sur un ensemble de données qui contient 360 modèles de visages 3D de 12 sujets. *Huang et al.* [222] présentent une méthode de reconnaissance faciale 3D basée sur les différences de statistiques d'images de profondeur. Un TEE de 12,4 % est atteint dans l'expérience de ROC III sur la base de données FRGC v.2.

Wang et al. [223, 224] emploient la SSDM (Signed Shape Difference Map) calculée entre deux visages 3D alignés comme une représentation pour la comparaison de la surface du visage. Sur la base des SSDMs, trois sortes de descripteurs (Haar, MB-LBP et *Gabor*) sont utilisés pour coder à la fois la similarité locale et les caractéristiques de modification entre les formes du visage. Les caractéristiques locales les plus discriminantes sont choisies de manière optimale par le renforcement et la formation comme des classificateurs faibles pour l'assemblage de trois classificateurs collectifs forts. Le taux de vérification rapporté par Roc III à 0,1 % TFA atteint 98,0 %.

Tang et al. [225] développent un algorithme de reconnaissance faciale 3D en utilisant une LBP (Local Binary Pattern) en variétés d'expression, extension de l'opérateur LBP, qui est largement utilisée dans l'analyse du visage 2D. Tout d'abord, pour décrire le visage humain avec plus de précision et réduire l'effet de sa distorsion locale, un système de division 3D du visage est proposé. Ensuite, pour chaque région du visage, l'histogramme statistique est utilisé pour résumer les détails du visage. Enfin, l'algorithme 3D de reconnaissance faciale proposée est testée sur les bases de données BJUT-3D et FRGC v2. Les auteurs ont obtenu des résultats prometteurs et ont conclu qu'il est possible d'appliquer la représentation de LBP sur la reconnaissance faciale 3D.

2.3.2.5 Les méthodes basées sur la fusion des caractéristiques multiples

Toutes ces caractéristiques locales mentionnées ci-dessus peuvent être combinées pour améliorer les performances de la reconnaissance faciale 3D et former ainsi la dernière catégorie: les méthodes basées sur la fusion des caractéristiques multiples.

Gordon [226] commence par la segmentation de visage basée sur la courbure. Ensuite, un ensemble de caractéristiques sont extraits décrivant à la fois la courbure et les propriétés de taille métrique du visage. Chaque visage devient donc un point dans un espace de fonction et la comparaison est faite par le plus proche voisin. Les expériences sont réalisées sur un ensemble de trois points de vue de chacun des 8 faces et un taux de reconnaissance élevé de 100% est rapporté dans l'ensemble de test.

Gokberk et al. [227] comparent cinq approches de reconnaissance faciale 3D. Ils comparent les méthodes basées sur la EGI, la correspondance par ICP, le profondeur de profil, la PCA et la LDA. Ils explorent également la fusion des résultats des cinq approches qui sont capables d'atteindre 99 % de taux de reconnaissance. À ce moment, ce travail est relativement nouveau dans la comparaison des performances des différents algorithmes de reconnaissance faciale 3D.

Dans [228], on a introduit à la fois une symétrie radiale et l'extraction de cinq caractéristiques sur les images de profondeur. Ces caractéristiques sont utilisées pour calculer la distance de Hausdorff.

Lee et al. [229] ont détecté manuellement huit points, sept courbes ainsi que des valeurs de distance et d'angle. La programmation dynamique est appliquée pour classer les courbes générées et ils ont obtenu un taux de reconnaissance de 95 % sur un ensemble de test de 5 individus, tandis que la SVM, utilisée pour classer les valeurs de distance et d'angle, a donné une performance de 96 % sur un ensemble de données de 100 individus.

Moreno et al. [230] ont calculé trente caractéristiques géométriques locales contenant la distance, l'angle, l'aire, les courbures, ...etc pour modéliser les surfaces faciales 3D. PCA et SVM sont utilisées pour la classification sur l'ensemble de données Gavab. Elles ont permis d'obtenir des résultats de 82,0 % et 90,16 %, lorsque les visages sont des vues frontales avec une expression neutre respectivement. Ces valeurs ont diminué à 76,2 % et 77,9 % dans le cas d'un geste, d'un changement de lumière et rotation du visage.

Li et Zhang [231] ont exploré l'utilisation de plusieurs attributs géométriques intrinsèques, y compris les angles, les distances géodésiques et les courbures pour la reconnaissance faciale 3D, où chaque face est représentée par un maillage triangulaire, prétraité pour posséder une connectivité uniforme. Ils forment alors des poids différents à appliquer à chaque attribut de la personne, ainsi que les pondérations utilisées pour combiner les attributs, afin de s'adapter aux variations d'expression. En utilisant l'approche de visage propre basée sur les poids appris et un classifieur de voisin le plus proche. Les résultats sur les bases de données Gavab et FRGC montrent que la fusion améliore la performance.

Li et al. [192] ont étudié la SRC (Sparse Representation Classifier) pour la reconnaissance faciale 3D en utilisant un ensemble de caractéristiques géométriques et ont réalisé un taux de reconnaissance de plus de 90 % des échantillons frontaux utilisant un ensemble de données composé de 120 personnes sélectionnées sur les bases de données Gavab et FRGC v2.

Guan et Zhang [232] combinent FSA (Facial Structural Angle) et LRM (Local Region Map) pour la reconnaissance faciale 3D. Les auteurs affirment que cette stratégie est robuste par rapport aux expressions faciales.

2.3.3 Autres approches

En plus des techniques globales et locales de reconnaissance du visage 3D, il existe aussi des approches visant à trouver une relation entre deux surfaces 3D du visage dans un espace de cartographie spéciale. Compte tenu de ces approches, nous les classons comme méthodes de paramétrisation de surface. Une paramétrisation d'une surface peut être considérée comme une mise en correspondance un-à-un de la surface dans un domaine approprié. En général, le domaine des

paramètres sera lui-même une surface et ainsi la construction d'une cartographie au moyen d'un paramétrage d'une surface à une autre. Ce paramétrage peut être résumé comme un problème d'optimisation sous certaines contraintes. Les surfaces faciales 3D qui sont représentées sous forme de maillages peuvent être mappées sur une image 2D régulière de sorte que des approches plus sophistiquées de traitement du signal peuvent être appliquées.

Abate et al. [233] transfèrent un maillage 3D à un plan 2D pour comparer deux visages représentés comme des surfaces polygonales 3D grâce à leurs cartes normales correspondantes (voir figure 2.23). Un réseau bidimensionnel stocke la courbure locale (maille normales) en tant que composantes RVB de pixels dans une image couleur. L'histogramme d'une carte de la différence résultant de la comparaison des cartes normales, pondérée par un masque d'expression, est utilisé comme caractéristique du visage pour la reconnaissance.

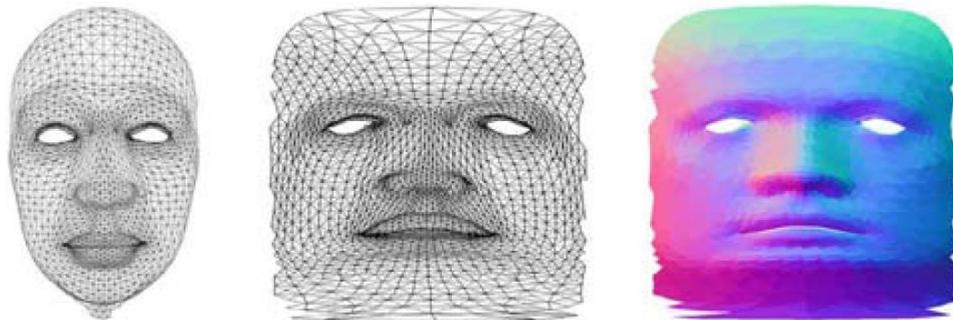


Figure 2.23: De maille géométrie ver la carte standard [233]

Bronstein et al. [234] utilisent la cartographie isométrique pour convertir les surfaces faciales expressives en un modèle standard, à savoir la forme canonique. On s'attend à ce que les modèles de visage avec différentes expressions d'un même sujet doivent généralement avoir la même forme canonique. Sur la base de cette idée, ils réalisent l'expérience sur un ensemble de tests contenant 220 modèles de visage de 30 sujets et rapportent un taux de reconnaissance de 100%. En outre, ils affirment également que cette méthode permet de distinguer des jumeaux identiques.

Pan et al. [235] explorent une profondeur d'image mappée pour la reconnaissance faciale 3D. Avec l'utilisation de la pointe du nez comme point central et un axe de symétrie pour l'alignement, les données de face sont mappées sur une plage de l'image circulaire. L'approche basée sur PCA est introduite pour la classification. La performance rapportée est 95% de taux de reconnaissance et un TEE= 2,8% dans un scénario de vérification sur l'ensemble de données FRGC v 1.

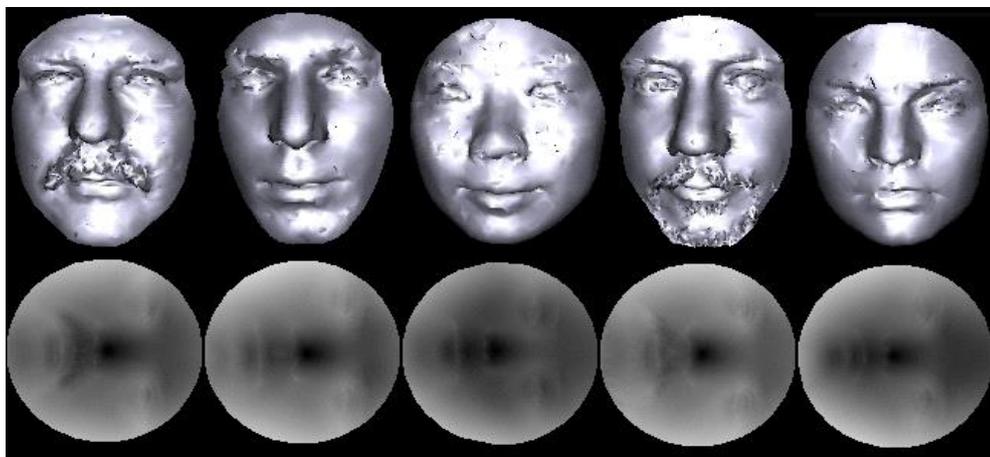


Figure 2.24: La première ligne: point de vue de rendu solide de ROI, et la deuxième ligne: image de profondeur relative mappée correspondant [234].

Passalis et al. [236] adaptent d'une part un modèle déformable prédéfini annoté manuellement d'une surface donnée du visage 3D de façon à pouvoir être paramétrée sur un plan 2D pour produire une image en couleurs, dont trois canaux contiennent les coordonnées des points de l'ensemble des sommets. La transformée en ondelettes est ensuite appliquée à l'image de la couleur et les coefficients correspondants sont adoptés en tant que vecteur de caractéristiques. Les expériences sur la base de données FRGC v2 montrent que cette méthode fonctionne beaucoup mieux que PCA et elle est également insensible aux variations d'expression et de genre.

2.3.4 Discussion

Après le développement deux dernières décennies, l'objectif actuel des techniques de reconnaissance faciale 3D est de répondre aux exigences de l'application en conditions réelles, c'est à dire de fournir des précisions élevées. Comme un sujet typique de la communauté de reconnaissance de formes, un système de reconnaissance faciale 3D contient l'étape d'extraction de caractéristiques et de classification. Sur la base de la taxonomie de la reconnaissance faciale 3D discutée ci-dessus, nous pouvons donner une conclusion, qui est similaire à celle du domaine de la reconnaissance de visage 2D, à savoir que les méthodes basées sur les caractéristiques locales surpassent généralement les méthodes globales. Les variations d'expression de visage est le principal facteur qui dégrade largement la précision d'un système de reconnaissance 3D. Afin d'améliorer la performance des méthodes basées sur les caractéristiques locales, une solution couramment adoptée consiste à combiner plusieurs représentations des formes locales de surfaces plus complètes [191, 224] ou la fusion des scores calculés à partir de différentes régions [238, 239, 220]. En outre, un ensemble d'apprentissage pour connaître la distribution des changements d'expression est également utile pour améliorer encore les résultats [204, 224].

Pendant ce temps, l'étape de prétraitement joue un rôle important dans la reconnaissance du visage. Comme les données 3D donnent des informations plus utiles que les données 2D, le prétraitement à l'aide des données 3D est plus pratique et précis. En règle générale, dans l'analyse de visage 3D, l'étape de prétraitement comporte deux parties principales, à savoir la détection des points de repère et l'enregistrement de la surface du visage. Quelques points de repère clés sont essentiels pour estimer les positions initiales des surfaces du visage ou localiser les traits du visage les plus complexes (par exemple, les courbes et les régions etc.).

2.4 Reconnaissance de visage 2D+3D

A la recherche d'une solution aux problèmes qui affectent de manière significative la reconnaissance faciale 2D, en particulier la sensibilité aux changements de pose, d'éclairage et de maquillage du visage, les chercheurs ont envisagé l'utilisation de modèles de visage 3D depuis 1989 et ont contribué par des améliorations significatives. Heureusement, la plupart des scanners de visage 3D (par exemple les scanners stéréoscopiques) capturent simultanément des images 2D sous des angles différents et les utilisent comme des données brutes dans leur méthode de calcul de la profondeur. Cette coexistence de modèles 2D et 3D suggère une possibilité d'amélioration de la performance de reconnaissance par une fusion de ces données biométriques indépendantes. Bien que la reconnaissance du visage 3D soit ancienne d'environ deux décennies, la reconnaissance multimodale de visage 2D+3D est au stade préliminaire et a eu une exposition limitée dans la littérature depuis l'année 2000. La majorité des algorithmes de cette catégorie utilise des méthodes classiques, comme les visages propres, pour créer des systèmes de reconnaissance indépendants 2D et 3D et pour combiner les scores de concordance par une stratégie de fusion.

Après l'approche purement 3D décrite dans [49], qui consiste à comparer les profils de modèles 3D après leurs recalage, *Beumier et al.* [240] proposent un nouveau schéma de reconnaissance multimodale par l'introduction d'une comparaison de profils 2D combinée avec la première approche.

La fusion est effectuée moyennant une somme pondérée de scores élémentaires. L'approche 2D+3D affiche un score plus élevé, TEE = 1.4%, par rapport à l'approche utilisant seulement les données 3D.

Dans le système de reconnaissance de visage proposé par Wang et al. (2002) [241], les descripteurs locaux des deux images de profondeur et d'intensité sont intégrés au niveau de données. Dans le domaine 2D, 10 points de référence détectés automatiquement sont décrites par les réponses des filtres de Gabor à différentes échelles et orientations. Les variations de forme dans une région autour de quatre points de repère détectés sur les images de profondeur sont capturés par des «signatures de points» [242]. Les vecteurs caractéristiques 2D et 3D sont initialement projetés à l'aide de la PCA. Les vecteurs caractéristiques normalisés de texture et de profondeur sont concaténés pour créer une représentation de visage. Deux méthodes de classification sont évaluées: une mesure de similarité et une SVM multi-class. Les résultats expérimentaux indiquent que le classificateur à base de SVM surpasse son homologue dans n'importe quelle combinaison de modalités 2D et 3D. Le taux de reconnaissance signalé pour la configuration 2D + 3D était supérieur à 90 % et il est démontré que la configuration multimodale a donné de meilleurs résultats que les configurations de modalités simples.

Bronstein et al. [243] proposent un schéma de décomposition utilisant la PCA multimodale 2D+3D. En 3D, c'est la transformation isométrique (canonical forms) qui a été utilisée, alors qu'en 2D il s'agit d'une approche de visages propres qui est appliquée sur les images de texture associées aux visages 3D appelées "attened textures". Une autre expérimentation réalisée par Chang et al. [59] consiste à appliquer une approche de visages propres sur des images de profondeur et des images d'intensité concernant 200 sujets. Les scores de reconnaissance obtenus sont égaux à 89 % pour l'approche 2D seule, 94 % pour l'approche 3D seule et 99 % pour l'approche hybride 2D+3D.

Lu et al. [244] proposent une méthode de reconnaissance de visage 2D+3D dans laquelle trois points caractéristiques sont localisés automatiquement à l'aide des signes de la courbure locale maximale et minimale, et une normalisation grossière est effectuée en suivant les lignes directrices de Weinstein [245]. L'alignement final est réalisé en minimisant la distance entre deux ensembles de points en utilisant une hybridation de deux algorithmes classiques ICP. Cet algorithme a été testé sur une base de données 73 images à partir de 10 individus. Le taux de reconnaissance obtenu est égal à 84%.

Tsalakanidou et al. [246] expérimentent dans l'approche de visages propres sur des images de profondeur et sur des images de couleur de 40 sujets de la base XM2VTS. Un taux de reconnaissance est obtenu par une approche multimodale 2D+3D autour de 99 % qui représente le taux le plus élevé que les taux obtenus par l'approche appliquée sur des données 2D seules ou 3D seules.

Dans [246], une description de l'approche, appelée "Visage", est donnée. La comparaison 3D est effectuée moyennant un appariement de graphes hiérarchiques, déjà utilisés dans leur système 2D commercialisé. Cette approche est plus rapide que les approches itératives de type ICP. La fusion est faite au niveau des scores de similarité. Les auteurs affirment un score de vérification de 93% correspondant à un taux de fausses acceptations (FAR) égal à 0,1%. Nous citons aussi l'approche "Geometrix", décrite dans [247] et qui fait appel à l'algorithme de comparaison 3D développé dans [248]. Un taux de vérification autour de 87 % est obtenu par la combinaison des deux approches correspondant à un taux de fausses acceptations égal à 0,1 %. Les évaluations des deux approches sont effectuées sur la base de données FRGC v2.

Wang et al. [249] considèrent l'intensité du visage $I(x, y, z)$ comme une fonction de coordonnées spatiales 3D et définissent les filtres de Gabor sphériques 3D (3D Spherical Gabor Filters, 3DSGF). Contrairement aux filtres traditionnels de Gabor, les filtres 3DSGFs sont à symétrie sphérique et n'ont pas donc la propriété de sélectivité directionnelle. Pour un point donné du visage, les réponses recueillies 3DSGFs de points voisins sont représentées collectivement par leur histogramme. Aucun algorithme de détection automatique de points clé n'est présenté. Les histogrammes de 3DSGFs sont calculés et comparés à l'aide de la distance "Least Trimmed Hausdorff" [250]. L'évaluation est effectuée sur une base de données de 960 paires d'images (2D et 3D) de 80 sujets capturées à l'aide de la technologie d'imagerie stéréo.

Hüsken et al. [246] appliquent la EBGM [74] séparément à l'image de profondeur et d'intensité pour créer un système parallèle avec 2DEBGM et 3DEBGM de reconnaissance de visage. Ces deux systèmes de reconnaissance sont combinés dans le niveau de comparaisons en additionnant les scores de similarité aux sorties de classificateurs. Semblable à 2DEBGM [74], l'étape de détection du point de repère est coûteuse en calcul puisque les zones de recherche de repères ne sont pas limitées. La détection du repère dans chaque modalité est indépendante et aucune tentative n'est faite pour les combiner. Comme 2DEBGM, les caractéristiques géométriques importantes, telles que les distances géodésiques entre repères, sont négligées dans ce projet de (2DEBGM + 3DEBGM).

Cook et al. [251] proposent un algorithme pour combiner les données de visage 2D et 3D pour améliorer la précision de reconnaissance. Dans cet algorithme, les images de profondeur et d'intensité sont convoluées avec les filtres *log-Gabor*. Chaque visage est divisé en 25 fenêtres carrées sur lesquelles une PCA est appliquée sur les réponses *log-Gabor*. Chaque fenêtre est considérée comme une observation indépendante où la distance de Mahalanobis est utilisée pour mesurer la similarité entre les fenêtres de test et de la galerie. Les distances de Mahalanobis résultant de la comparaison sont recueillies dans un vecteur de caractéristiques utilisé par une machine à vecteurs de support linéaire (SVM). Les auteurs montrent que les caractéristiques 3D *log-Gabor* obtiennent des résultats nettement meilleurs que ceux de 2D *log-Gabors*.

Mpiperis et al. [252] proposent un algorithme de reconnaissance de visage 2D+3D où les expressions du visage (images de profondeur et d'intensité) sont compensées par le paramétrage polaire géodésique de la surface du visage. Ils ont testé leur algorithme sur une base de données de 1500 images de 100 sujets et un TEE de l'ordre de 4,9 était signalé à 15,4 %.

Xu et al. [253] proposent un modèle de reconnaissance de visage en fonction des réponses d'ondelettes de *Gabor* extraites sur les images de profondeur et d'intensité à 5 échelles et 8 orientations. Tout d'abord, les images de profondeur et d'intensité sont alignées et normalisées par rapport à un modèle de visage générique. Les vecteurs caractéristiques de très grandes dimensions pour chaque modalité (quatre-vingts fois le nombre de pixels dans l'image) sont calculés. Les caractéristiques 2D et 3D de *Gabor* sont converties en sous-espaces de dimension, puis fusionnées au niveau de leur caractéristique en utilisant une combinaison de LDA et Adaboost [254]. La manipulation des caractéristiques de très grandes dimensions est également coûteux en calcul (par exemple dans [253], au moins 80 LDA sont appliquées dans le premier niveau de la hiérarchie).

Al-Osaimi et al. [255] proposent la fusion optimisée au niveau de données (pixel-level fusion) 2D et 3D pour la reconnaissance faciale. Les fonctions de fusion sont objectivement optimisées pour modéliser les variations d'expression et d'éclairage de sous-espaces linéaires invariants de reconnaissance de visage. Les paramètres des fonctions adjacentes sont contraints de varier en douceur pour une régularisation numérique efficace. En plus de l'optimisation spatiale, plusieurs modèles de fusion non linéaire sont combinés pour renforcer leurs capacités d'apprentissage. Les expériences sur la base de données FRGC v2 montrent que l'optimisation spatiale et la combinaison de plusieurs de ces fonctions de fusion améliorent systématiquement les performances; ce qui est, pour la première fois, plus élevé que la fusion au niveau de scores dans un dispositif expérimental similaire.

Wang et al. [256] utilisent la méthode "Adaboost" pour l'extraction de la région faciale. Puis l'extraction de caractéristiques 2D et 3D du visage est faite par DTCWT (Dual-tree Complex Wavelet Transform). Les quatre niveaux des composantes hautes fréquences de l'image de texture 2D et l'image de la profondeur 3D sont obtenus respectivement, puis l'analyse discriminante linéaire (LDA) est utilisée pour obtenir les vecteurs des caractéristiques. Cosinus distance est développée pour la création de deux matrices de similarité. Enfin un résultat de fusion est établi par les deux matrices de similarité. Le taux de vérification sur un TFA de 0,1 % est de 97,6 % sur le protocole "All vs All" (tous contre tous).

En conclusion, la plupart des approches multimodales (2D+3D) confirment l'intérêt de la fusion des deux modalités en vue d'obtenir des taux de reconnaissances et d'authentification plus élevés que les approches exploitant les données 2D seules ou 3D seules.

2.5 Conclusion

Dans la première partie de ce chapitre, nous avons minutieusement examiné plusieurs techniques qui sont considérées pour la reconnaissance de visage 2D en fonction de l'image d'intensité, tandis que dans la seconde partie, nous avons examiné les approches de représentation 3D basée sur la reconnaissance faciale 3D. Enfin dans la troisième partie nous avons discuté les modèles de reconnaissance faciale multimodales (2D+3D).

Dans la modalité 2D, presque toutes les méthodes récentes affirment être en mesure de fournir des taux de reconnaissance satisfaisants, mais seulement quand elles sont évaluées sur des bases de données standard. En fait, certains des défis fondamentaux, à savoir les variations d'éclairage et de pose, n'ont pas été complètement résolus. La modalité 3D a la capacité d'améliorer les performances de reconnaissance faciale 2D par rapport à ces défis. De nombreuses techniques ont déjà prouvé leur capacité dans certains cas de validation. Cependant, plusieurs autres problèmes difficiles ont encore besoin d'être dûment pris en compte, comme la précision, la vitesse d'acquisition des données, la précision de détection des points clé et la lenteur de l'enregistrement de la surface. En outre, la modalité 3D a tendance à être plus sensible aux variations d'expression du visage par rapport à celle basée sur des images 2D.

En conséquence, plusieurs chercheurs ont proposé de meilleures solutions pour améliorer la performance dans chaque modalité. Nous pouvons également continuer ce travail sur la fusion des deux modalités (2D et 3D) afin que la robustesse des systèmes de reconnaissance faciale puisse être améliorée.

3.1 Introduction

La reconnaissance faciale est une tâche que les humains effectuent naturellement et sans effort dans leur vie quotidienne. La grande disponibilité d'ordinateurs puissants et peu onéreux ainsi que des systèmes informatiques embarqués ont suscité un énorme intérêt dans le traitement automatique des images au sein de nombreuses applications, incluant l'identification biométrique, la surveillance, l'interaction homme-machine et la gestion de données multimédia. La reconnaissance faciale, en tant que technologie biométrique de base, a pris une part de plus en plus importante dans le domaine de la recherche. Ceci est dû aux avancées rapides dans les technologies. De nombreuses techniques et architectures 2D ont été développées ces dernières années. Nous les avons passées en revue dans le chapitre 2. Nous allons donc présenter dans ce chapitre notre système monomodal et multi-algorithmes de vérification de visage 2D. Dans la partie de vérification de visage 2D monomodale ou uni-modale, nous présentons tout d'abord notre modèle de vérification de visage, puis nous détaillons les quatre étapes principales de ce modèle. Enfin l'étude expérimentale de l'authentification de visages sur la base de données XM2VTS est réalisée par la comparaison des algorithmes de réduction d'espace suivie de l'étude de l'extraction des caractéristiques par les ondelettes de Gabor. Dans la partie de vérification de visage 2D multi-algorithmes, les limitations des systèmes biométriques lorsqu'ils n'utilisent qu'une seule modalité biométrique sont passées en revue. Nous présentons ensuite les différentes multi-modalités possibles. L'état de l'art de la fusion de scores et leurs méthodes de fusion sont présentés. Nous terminons notre chapitre par les résultats expérimentaux obtenus.

3.2 Vérification uni-modale de visage 2D

Le principe de fonctionnement de base de notre système de vérification faciale (figure 3.1) peut être résumé en **quatre étapes**: prétraitement, extraction de caractéristiques, réduction d'espace et comparaison.

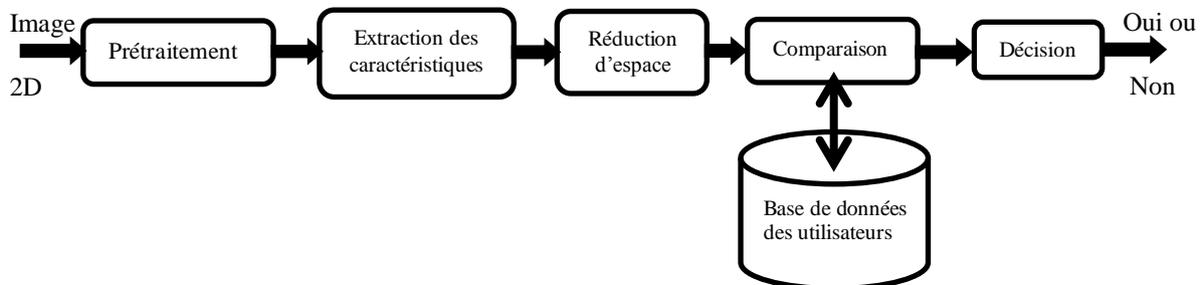


Figure 3.1: Principe de fonctionnement de notre système de vérification de visage.

3.2.1 Prétraitement d'image

Le prétraitement est une phase importante dans le processus global d'authentification. C'est une méthode simple qui augmente en général les performances du système. Elle permet souvent une première réduction des données et elle atténue les effets d'une différence de conditions lors des prises de vues. Dans notre travail, nous utilisons trois phases de prétraitement: le découpage de l'image, la décimation et la conversion de la couleur dans l'espace de couleur TSV.

3.2.1.1 Découpage

Le découpage de l'image consiste à conserver les maximums des variations intrinsèques du visage, et à supprimer les autres informations comme l'arrière plan, les cheveux, les cols de chemise, les oreilles...etc. afin d'améliorer les performances du système d'authentification. Une fenêtre

rectangulaire de taille (161 × 121), centrée autour des caractéristiques les plus stables liées aux yeux, aux sourcils, au nez et à la bouche, a été utilisée. La figure 3.2 montre l'utilisation de cette fenêtre.

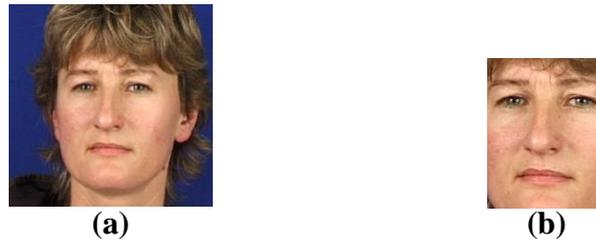


Figure 3.2: Image de visage (a) avant et (b) après découpage

3.2.1.2 Décimation

La décimation consiste à ne prendre qu'un pixel sur quatre. Cela réduit bien entendu la résolution des images. L'image de visage passera ainsi d'une dimension $256 \times 256 = 65536$ pixels à une dimension de $81 \times 81 = 4941$ pixels après découpage et décimation, comme il est illustré sur la figure 3.3.

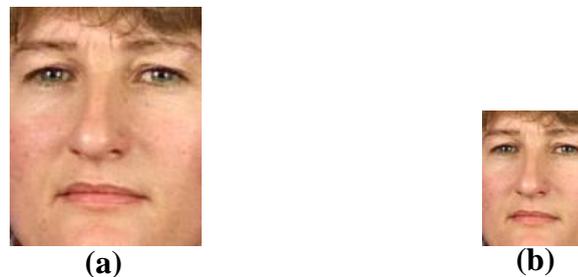


Figure 3.3: Image de visage (a) avant (b) après décimation.

3.2.1.3 Composante S de l'espace de couleur TSV de l'image

Les systèmes d'authentification de visages utilisent souvent la représentation en niveaux de gris de l'image de visage comme caractéristique d'entrée. Mais lorsque les images sont en couleurs, il est avantageux d'utiliser cette information supplémentaire pour améliorer les performances de ce système d'authentification. Nous utilisons donc l'espace de couleur **TSV** (**Teinte, Saturation, Valeur**) ou **HSV** (**Hue, Saturation, Value**). Ce choix est justifié dans [257] où l'on confirme que c'est le système le plus utilisé dans la littérature. La quantification de la couleur de notre base de données XM2VTS est faite dans le système **RVB** (**Rouge, Vert et Bleu**). Le passage du système de couleurs **RVB** au système **TSV** se fait par une fonction de transfert non linéaire. La figure 3.4 présente les deux systèmes colorimétriques **RVB** et **TSV**.

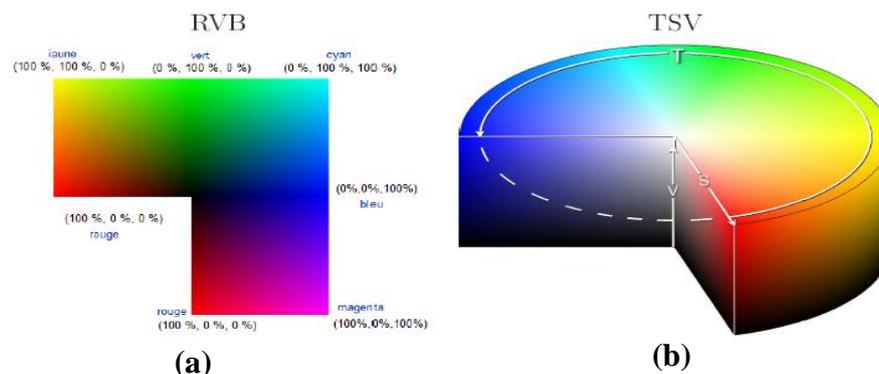


Figure 3.4: Différents systèmes colorimétriques: (a) plan du système de couleurs **RVB**, (b) cylindre du système de couleurs **TSV**.

Soient min et max , respectivement, les valeurs minimale et maximale des composantes R , V et B ($min = minimum(R, V, B)$ et $max = maximum(R, V, B)$) et soient x et y les coordonnées d'un pixel de l'image. Les transformations des espaces RVB vers TSV sont données par les équations suivantes [257].

$$T(x, y) = \begin{cases} 0, & si \quad min = max \\ 60 * \frac{V-B}{max-min}, & si \quad max = R \text{ et } V \geq B \\ 60 * \frac{V-B}{max-min} + 360, & si \quad max = R \text{ et } V < B \\ 60 * \frac{B-R}{max-min} + 120, & si \quad max = V \\ 60 * \frac{R-V}{max-min} + 240, & si \quad max = B \end{cases} \quad (3.1)$$

$$s(x, y) = \begin{cases} 0, & si \quad max = 0 \\ 1 - \frac{min}{max}, & si \quad non \end{cases} \quad (3.2)$$

$$V(x, y) = max \quad (3.3)$$

Pour plus de détails sur cette transformation, on peut se référer à [258]. Dans [59], on a étudié les espaces de couleurs sur la base de données XM2VTS pour les deux protocoles de *Lausanne* et on a trouvé que la composante S donne le meilleur résultat pour la tâche de vérification d'identité de visages. Partant de ce résultat, nous utilisons donc la composante S de l'espace de couleur TSV . La figure 3.5 présente la conversion de couleur et l'extraction de la composante S du visage.

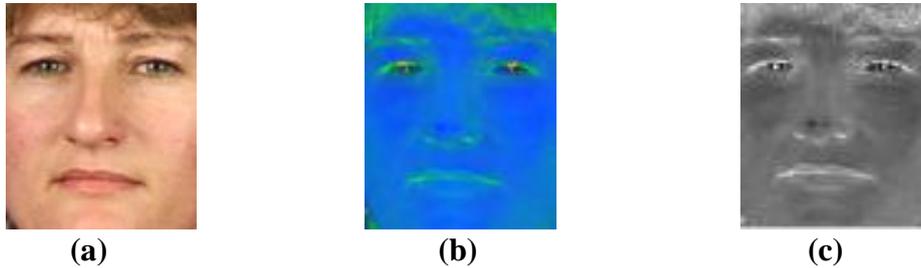


Figure 3.5: (a) Image découpée et décimée, (b) Image en système TSV, (c) Composante S .

3.2.2 Extraction des caractéristiques de visage par les ondelettes de *Gabor*

Les ondelettes de *Gabor* sont connues comme le moyen d'analyse espace-fréquence qui minimise l'incertitude de *Heisenberg*, à savoir si on est précis dans l'analyse de l'information dans l'espace du pixel on le sera moins dans l'espace fréquentiel et vice-versa. Cette caractéristique fait des ondelettes de *Gabor* un moyen puissant pour l'analyse de la texture. Elles ont été développées au début pour l'analyse des signaux à une dimension. Leur expression générale est :

$$W(t, t_0, \omega) = e^{-\sigma(t-t_0)^2} e^{-i\omega(t-t_0)} \quad (3.4)$$

Où ω est la fréquence et $i^2 = -1$. Cette ondelette est en fait une gaussienne modulée par une fonction sinusoïdale. La généralisation de cette fonction en **2D** donne une fonction gaussienne à deux dimensions modulée par un harmonique. Les ondelettes de *Gabor* 2D ont la caractéristique d'être modulables par différents paramètres qui contrôlent leurs orientations (θ), fréquences (inversement proportionnelle à λ), phase de la sinusoïde (ϕ), le support temporel (σ) et l'enveloppe de la gaussienne (γ) [4]. L'équation générale d'une ondelette de *Gabor* en 2D est :

$$W(x, y, \theta, \lambda, \varphi, \sigma, \gamma) = e^{-\frac{x'^2 + y'^2}{2\sigma^2}} e^{i\left(\frac{x'}{\lambda} + \varphi\right)} \quad (3.5)$$

Où :

$$x' = x \cos(\theta) + y \sin(\theta) \text{ et } y' = -x \sin(\theta) + y \cos(\theta) \quad (3.6)$$

3.2.3 Algorithmes de réduction d'espace de données

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 dimensions. Après normalisation de l'image et mise à l'échelle vers une taille fixe (ex : $m*n$), l'image du visage peut être considérée comme un vecteur dans un espace multidimensionnel R^{mn} . Le nombre de points constituant cet espace devient rapidement très grand, même pour les images de petite dimension.

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 des 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.

En 1994, *Ruderman* a démontré que les images naturelles possèdent une grande redondance statistique [260]. En 1996, *Penev* [261] a démontré que dans le cas précis des images normalisées des visages, cette redondance statistique est très 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 dimensionnalité 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 [262].

3.2.3.1 Analyse en Composantes Principales (ACP)

L'analyse en composantes principales (ACP) est une méthode mathématique qui peut être utilisée pour simplifier un ensemble de données, en réduisant sa dimension et pour représenter efficacement les images de visages. Ainsi, au lieu de stocker un vecteur image de taille N , on stocke un vecteur de taille $R \leq$ nombre d'images d'apprentissage P . Les étapes principales pour calculer la matrice de projection (U_{PCA}) de ACP sont :

- chaque image est représentée par un vecteur de taille N .

$$x^i = [x_1^i, x_2^i, \dots, x_N^i]^T \quad (3.7)$$

- toutes les images sont centrées en soustrayant l'image moyenne de chaque vecteur image,

$$\bar{x}^i = x^i - m, \quad m = \frac{1}{P} \sum_{i=1}^P x^i \quad (3.8)$$

m : moyenne de toutes les images d'apprentissage.

Ces vecteurs sont combinés, côte-à-côte, pour créer une matrice \bar{X} de données d'apprentissage de taille $N \times P$ (où P est le nombre d'images d'apprentissage, N la taille du vecteur image).

Le calcul de la matrice de covariance (C) qui est définie par la formule :

$$C = \bar{X} \bar{X}^T \quad (3.9)$$

On calcule ensuite les vecteurs propres V et les valeurs propres D de la matrice C . Les valeurs propres de la matrice C représentent le taux de variation le long de l'axe du vecteur propre associé. La matrice de transformation de la ACP est donc les R premiers vecteurs propres ordonnés par ordre décroissant des valeurs propres correspondantes (U_{ACP}).

- Inconvénient

L'ACP ne tient pas compte de l'aspect (discrimination) des classes et on ne devrait pas s'attendre ainsi à une performance optimale pour des tâches telles que l'authentification de visage en utilisant la ACP pour le codage.

- Solution

Une solution qui a été proposée pour résoudre ce problème est d'employer l'analyse discriminante linéaire de Fisher dans le but même de réaliser une séparabilité élevée entre les différentes classes dans lesquelles la classification est intéressante.

3.2.3.2 Analyse Discriminante Linéaire (LDA)

L'analyse discriminante linéaire (Linear Discriminant Analysis (LDA)) est utilisée pour trouver la combinaison linéaire des caractéristiques qui séparent le mieux les classes d'objet ou d'événement. Les combinaisons résultantes peuvent être employées comme classificateur linéaire, ou généralement dans la réduction de caractéristiques.

La LDA est une technique qui cherche les directions qui sont efficaces pour la discrimination entre les données. La figure 3.6 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 l'ACP.

Concrètement, pour tous les échantillons de toutes les classes, on définit deux mesures. La première mesure est la matrice de dispersion intra-classes S_w ("within-class scatter matrix") qui est définie par :

$$S_w = \sum_{i=1}^L \frac{1}{n_i} \sum_{j=1}^{n_i} (x_j^i - m^i)(x_j^i - m^i)^T \quad (3.10)$$

Avec x_j^i le $j^{\text{ème}}$ échantillon de la classe i , m^i la moyenne de la classe i , L le nombre de classes et n_i le nombre d'échantillons de la classe i . La deuxième mesure est la matrice de dispersion inter-classes S_b ("between-class scatter matrix") qui est définie par:

$$S_b = \sum_{i=1}^L (m^i - \bar{m})(m^i - \bar{m})^T \quad (3.11)$$

Avec \bar{m} la moyenne de tous les échantillons.

Le but de LDA est de maximiser les distances inter-classes tout en minimisant les distances intra-classes, ce qui revient à retrouver la matrice de transformation U_{LDA} qui maximise le critère:

$$W_{opt} = \arg \max_W \left(\frac{|W^T S_b W|}{|W^T S_w W|} \right) \quad (3.12)$$

Ce problème est ramené à un problème de recherche des vecteurs propres de la matrice $S_w^{-1} S_b$. La matrice de transformation de LDA est les m premiers vecteurs propres sont donc ordonnés par ordre décroissant des valeurs propres correspondantes (U_{LDA}).

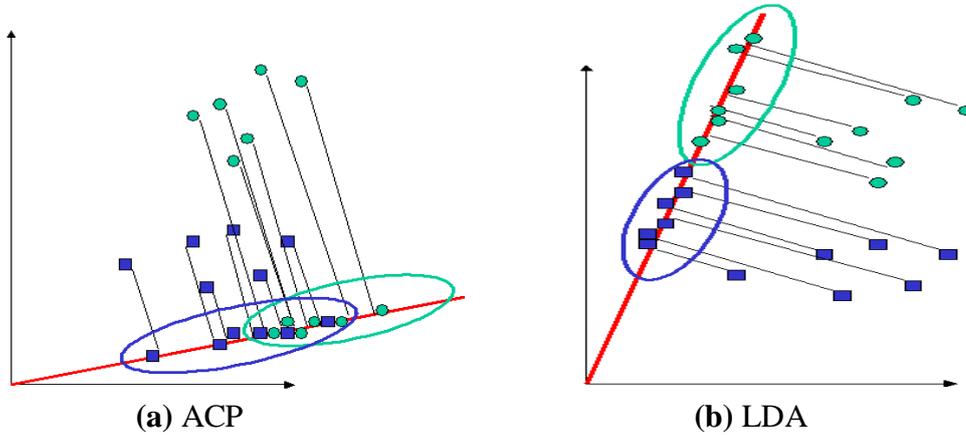


Figure 3.6: Comparaison entre les projections de deux classes de points ("classe 1" et "classe 2") sur les axes principaux construits par (a) ACP et par (b) LDA.

- **Inconvénient**

LDA exige un grand nombre d'échantillons de l'ensemble d'apprentissage pour la bonne généralisation. Quand un tel besoin n'est pas obtenu, la LDA crée un problème de sur-ajustement aux données d'apprentissage et ceci se prête mal aux nouvelles données de test [262, 104, 96].

- **Solution**

Le modèle discriminant Linéaire Amélioré de *Fisher* (Enhanced *Fisher* Model (EFM)).

3.2.3.3 Modèle discriminant linéaire amélioré de *Fisher*

EFM améliore la capacité de généralisation de la LDA en décomposant la procédure LDA en diagonalisation simultanée des deux matrices de dispersion intra-classe et inter-classe [263]. La diagonalisation simultanée est une étape équivalente à deux opérations comme l'a souligné *Fukunaga* [19]: blanchiment de la matrice de dispersion intra-classe et l'application de la ACP sur la matrice de dispersion inter-classe en utilisant les données transformées. Durant l'opération de blanchiment de la matrice de dispersion intra-classe apparaissent dans le dénominateur de la séparabilité de petites valeurs propres qui tendent à capturer du bruit [263]. Pour atteindre des performances améliorées, l'EFM préserve un équilibre approprié entre la sélection des valeurs propres (correspondant à la composante principale de l'espace de l'image originale) qui tiennent compte de la plupart de l'énergie spectrale des données brutes, c.-à-d., la représentation adéquate et l'exigence que les valeurs propres de la matrice de dispersion intra-classe (de l'espace ACP réduit) ne sont pas trop petites, c.-à-d., meilleure généralisation.

Le choix du rang des composantes principales (m), pour la réduction de la dimension, prend en compte l'ordre de grandeur de l'énergie spectrale. Les valeurs propres de la matrice de covariance fournissent un bon indicateur pour répondre au critère de l'énergie. Il faut ensuite calculer les valeurs

propres de la matrice de dispersion intra-classe dans l'espace ACP réduit pour faciliter le choix du rang des composantes principales de sorte que l'exigence de grandeur est respectée. A cette fin, on effectue la LDA par étapes comme décrit ci-dessous. En particulier, ces étapes LDA permettent de trouver les valeurs propres et les vecteurs propres de $S_W^{-1}S_b$ comme résultat de la diagonalisation simultanée de S_W et S_b .

Alors l'étape de EFM sont présentées comme suit :

- Blanchissons d'abord la matrice de dispersion intra-classe :

$$S_W V = V \Lambda \quad \text{et} \quad V^T V = I \quad (3.13)$$

$$A^{-1/2} V^T S_W V A^{-1/2} = I \quad (3.14)$$

Où $V, \Lambda \in \mathbb{R}^{m \times m}$ sont la matrice des vecteurs propres et la matrice diagonale des valeurs propres de S_W respectivement.

L' EFM diagonalise donc en premier lieu la matrice de dispersion intra-classe S_W (3.10).

- En second lieu EFM procède à calculer la nouvelle matrice de dispersion inter-classe comme suit:

$$K_b = A^{-1/2} V^T S_b V A^{-1/2} \quad (3.15)$$

- Diagonalisons maintenant la nouvelle matrice de dispersion inter-classe K_b :

$$K_b V_b = V_b \Lambda_b \quad \text{et} \quad V_b^T V_b = I \quad (3.16)$$

Où $V_b, \Lambda_b \in \mathbb{R}^{m \times m}$ sont la matrice des vecteurs propres et la matrice diagonale des valeurs propres de K_b respectivement.

- La matrice de transformation globale de l'EFM est définie comme suit [19]:

$$U_{EFM} = V A^{-1/2} V_b \quad (1.17)$$

3.2.3.4 LDA directe (DLDA)

Une nouvelle approche, appelée la DLDA, a été proposée dans [264] 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 :

$$V^T S_b V = \Lambda \quad (3.18)$$

Avec: $V^T V = I$.

Le problème est alors ramené à la résolution d'un problème aux 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) :

$$Y^T S_b Y = D_b > 0 \quad (3.19)$$

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

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

$$\left(YD_b^{-\frac{1}{2}}\right)^T S_b \left(YD_b^{-\frac{1}{2}}\right) = D_b \quad (3.20)$$

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

$$Z^T S_b Z = D_b \quad (3.21)$$

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 :

$$U^T Z^T S_w Z U = D_w \quad (3.22)$$

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* :

$$AS_w A^T = D_w \quad , \quad AS_w A^T = I \quad (3.23)$$

En posant $U_{DLDA} = D_w^{-1/2} A$, nous obtenons la matrice de projection qui répond au critère de *Fisher* [257].

3.2.4. Comparaison

Lorsqu'on souhaite comparer deux vecteurs de caractéristiques réduits issus du module de réduction de données d'un système biométrique, on peut soit effectuer une mesure de similarité (ressemblance), soit une mesure de distance (divergence).

Bien que la distance euclidienne soit optimale dans la théorie, les diverses expériences ont trouvé que la distance euclidienne est surpassée par d'autres distances. Une d'elles est le cosinus [265] qui est définie par:

$$S(A, B) = \frac{A^T B}{\|A\| \|B\|} \quad (3.24)$$

Cette fonction calcule simplement le cosinus de l'angle entre les deux vecteurs caractéristiques A et B . Une valeur élevée de corrélation normalisée correspond à une bonne similarité entre les deux vecteurs.

3.2.5 Résultats

3.2.5.1 Base de données XM2VTS

La base de données XM2VTS (Extended Multi Modal Verification for Teleservices and Security applications) [266] est une base multimodale développée au sein du projet européen ACTS. Elle offre des enregistrements synchronisés des photos de visages prises de face et de profil et des paroles de 295 personnes des deux sexes hommes et femmes de différents âges. Elle est utilisée pour la vérification d'identité. Pour chaque personne huit prises ont été effectuées en quatre sessions distribuées pendant cinq mois afin de prendre en compte le changement d'apparence des personnes selon plusieurs facteurs (présence et/ou absence de lunettes, barbe, changement de coiffure, pose, etc.). Par contre, seules les expressions neutres ont été considérées et chaque session est composée de deux enregistrements, un pour les séquences de parole et l'autre pour les séquences vidéo de la tête. Les vidéos et photos sont en couleur de haute résolution (format PPM). La taille est de 256×256 pixels pour les images et sont de très bonne qualité codées sur 24 bits dans l'espace RGB. Cela permet de travailler en niveaux de gris ou en couleur.



Figure 3.7: Exemples des images de la base de données XM2VTS.



Figure 3.8: Exemples des images d'une personne de la base de données XM2VTS

Le choix de XM2VTS est justifié par sa grande taille, avec 295 personnes et 2360 images au total et sa popularité puisqu'elle est devenue une norme dans la communauté biométrique audiovisuelle de vérification multimodale d'identité.

Pour la tâche de vérification, un protocole standard d'estimation de performances a été développé et appelé «*Lausanne protocol splits randomly*». Il existe deux configurations différentes LP I et LP II définies dans [14]. Nous utiliserons la configuration I puisqu'elle est la plus difficile pour la vérification de visage. Elle consiste à séparer les individus en deux classes, client et imposteur. Le groupe client contient 200 sujets, alors que le groupe imposteur est divisé en 25 imposteurs pour l'évaluation et 70 imposteurs pour le test. La répartition des photos dans les différents ensembles est présentée dans le tableau 3.1, et les figures 3.7, 3.8 présentent des images de face de la base de données XM2VTS.

Tableau 3.1: Répartition des photos dans les différents ensembles (LP I).

Ensemble	Clients	Imposteurs
Apprentissage	600 images (3 par personne)	0 images
Evaluation	600 images (3 par personne)	200 images (8 par personne)
Test	400 images (2 par personne)	560 images (8 par personne)

3.2.5.2 Comparaison des algorithmes de réduction de l'espace de données

Dans le visage de la même personne, deux images très semblables (à l'œil humain) pourraient être extrêmement différentes si elles sont comparées pixel par pixel. Il est donc nécessaire d'extraire les caractéristiques appropriées et discriminantes à partir des images du visage et de comparer ces caractéristiques au lieu de comparer des images approximatives. Naturellement, plus les caractéristiques sont discriminantes, plus l'authentification sera facile.

Nous avons étudié 4 algorithmes de réduction d'espace de données pour choisir le meilleur algorithme. Nous utilisons plusieurs algorithmes de réduction d'espace. On peut légitimement se poser la question: quel algorithme choisir ? Pour répondre à cette question, nous avons effectué nos expériences sur tous ces algorithmes pour le choix du meilleur algorithme.

La figure 3.9 nous montre que la méthode EFM est plus performante que la LDA suivie de la DLDA suivie de L'ACP en employant un nombre restreint de caractéristiques (vecteurs propres) dans notre base de données XM2VTS.

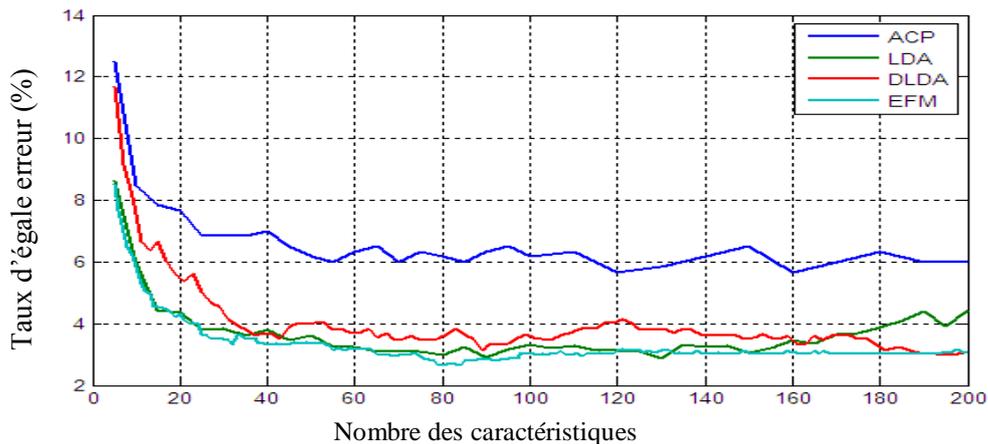


Figure 3.9: Comparaison des TEE pour ACP, LDA, EFM et DLDA dans l'ensemble d'évaluation.

Le tableau 3.2 nous montre bien que le meilleur résultat est obtenu avec EFM (associée avec ACP) avec un TEE=2.66 % et un TV=94.33 % en utilisant seulement 80 paramètres caractéristiques. Nous remarquons aussi que les TV et TEE obtenus par la DLDA et la LDA sont assez proches. Alors que le mauvais résultat est obtenu par ACP appliquée seule. C'est un résultat bien évident puisque

cette méthode est gourmande en quantité de données et n'assure pas une discrimination optimale. L'algorithme de réduction d'espace de données retenu reste l'EFM.

Tableau 3.2: Performances des méthodes ACP, LDA, EFM, DLDA.

Méthodes	Ensemble évaluation		Ensemble de test		Nombre de caractéristiques
	TEE (%)	TFR (%)	TFA (%)	TV (%)	
ACP	5.7	2.7	8.1	89.2	200
ACP+LDA	2.83	2	4.12	93.88	90
ACP+EFM	2.66	2	3.66	94.33	80
ACP+DLDA	3	2.1	4.3	93.6	195

3.2.5.3 Extraction des caractéristiques par les ondelettes de Gabor

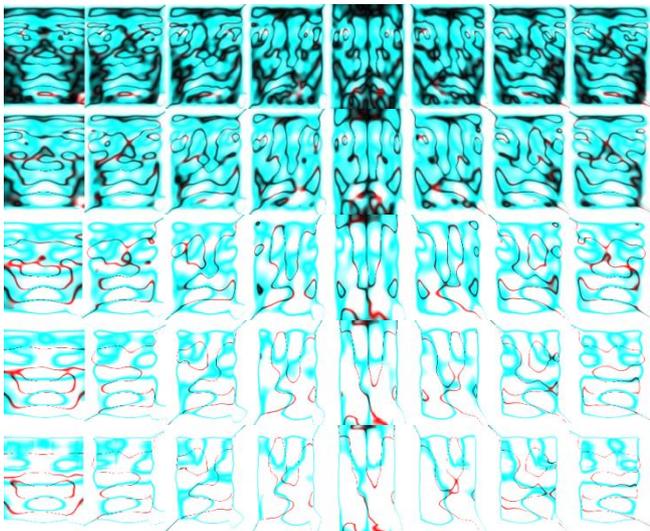
La méthode de reconnaissance que nous étudions dans cette partie repose principalement sur l'extraction des caractéristiques du visage par filtrage de *Gabor* 2D. Les filtres de *Gabor* sont connus comme un moyen d'analyse espace-fréquence très robuste. Cette spécificité a fait des filtres de *Gabor* un moyen puissant d'analyse de textures et de classification. Les filtres de *Gabor* analysent la texture d'un objet suivant différentes résolutions et différents angles dans le domaine spatial.

La représentation de *Gabor* d'une image de visage est obtenue par la convolution de l'image avec la famille des filtres de *Gabor*. Cette convolution est 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 .

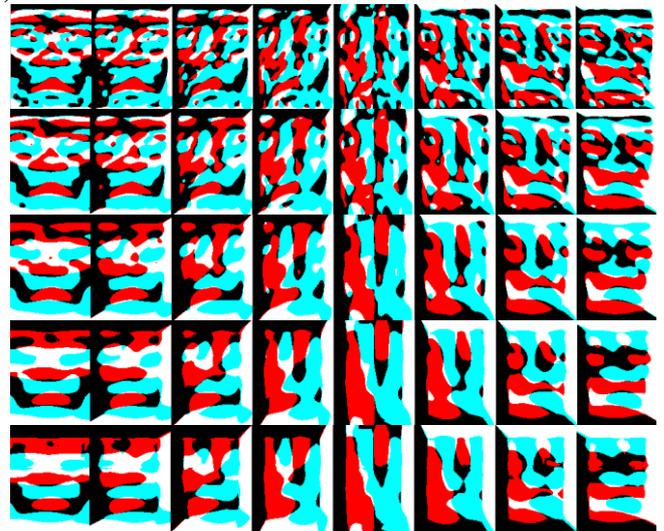
La famille de filtres de *Gabor* est alors caractérisée par un certain nombre de résolutions ou fréquences et orientations qui seront appelées "caractéristiques" par la suite. Comme on peut le remarquer dans l'expression 3.5, les filtres de *Gabor* ont une forme complexe qu'il est possible d'exploiter. En effet, étant en quadrature de phase, il est important d'utiliser les informations données par les parties réelle et imaginaire des coefficients de *Gabor*.



(a)



(b)



(c)

Figure 3.10: Résultats de la convolution d'une image de visage avec une famille de 40 filtres de *Gabor*
 (a) Image en espace de couleur TSV, l'ensemble (b) amplitudes et (c) les phases.

Deux choix triviaux s'offrent à nous: l'étude de l'amplitude et l'étude de la phase de *Gabor*. Dans [267], *Oppenheim* a démontré que la phase des analyses multi-résolutions est plus informative que les amplitudes qui peuvent souffrir des problèmes d'illumination.

La figure 3.10 présente les amplitudes et les phases de la convolution de l'image de visage en l'espace de couleur TSV par quarante filtres de *Gabor*.

La majorité des approches se basant sur le filtrage de *Gabor* pour la reconnaissance de visages (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 [268, 269, 270]. 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 [259].

- Influence des caractéristiques de la famille des filtres de *Gabor* sur les performances de l'authentification de visages

Nous commencerons par étudier l'influence des caractéristiques des familles des filtres de *Gabor* sur les performances de l'authentification pour en déduire le choix optimal. La représentation de l'image considérée est l'amplitude des réponses des filtres de *Gabor*. L'image d'entrée est une image de visage coupée et décimée. Nous considérons la composante *S* de l'image couleur dans l'espace TSV. L'algorithme de réduction d'espace utilisé est ACP + EFM appliquée à l'amplitude des réponses des filtres de *Gabor* et la mesure de similarité utilisée est le *cosinus*. Le tableau 3.3 présente les résultats de l'authentification en terme taux d'égale erreur (TEE) dans l'ensemble d'évaluation pour différents niveaux de résolution et d'orientations de quarante filtres de *Gabor* et un nombre de caractéristiques dans la phase de réduction d'espace par EFM égale $N_p=100$.

Tableau 3.3: Taux d'égale erreur TEE dans l'ensemble d'évaluation pour différents niveaux de résolutions et d'orientations pour les 40 filtres.

Résolution (λ)	Orientations des filtres (θ)							
	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$	$5\pi/8$	$3\pi/4$	$7\pi/8$
4	9.28	10.1	8.13	8.02	8.01	8.04	8.63	7.3
$4\sqrt{2}$	8.33	9.01	7.54	9.61	5.35	7.95	7.2	8.5
8	9.31	7.34	8.7	5.7	7.85	5.13	8.02	8.17
$8\sqrt{2}$	9.54	8.64	7.31	9.36	10.19	8.3	7.07	7.54
16	9.17	8.48	8.65	9.18	9.18	8.64	7.84	7.62

Dans ce tableau nous constatons que le meilleur TEE= 5.13% avec résolution ($\lambda=8$) et orientation ($\theta = 5\pi/8$). Ce n'est pas encore un bon résultat ! Pour cela nous nous intéressons aux phases des filtres de *Gabor*.

- Utilisation de la phase de *Gabor* pour l'authentification de visages

Lorsque nous considérons une image du visage, 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 $Real(IG_{r,o}) \cong 0$ et $Im(IG_{r,o}) \cong 0$. 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 sélectionnons les régions informatives par seuillage de l'amplitude à chaque point d'analyse.

$$P(IG_{r,o}(x, y)) = \begin{cases} \arctan\left(\frac{Im(IG_{r,o}(x,y))}{Real(IG_{r,o}(x,y))}\right) & \text{si } M(IG_{r,o})(x, y) \geq Th \\ 0 & \text{si } M(IG_{r,o})(x, y) < Th \end{cases} \quad (3.25)$$

$$M(IG_{r,o})(x, y) = \sqrt{(real(IG_{r,o}(x, y)))^2 + (Im(IG_{r,o}(x, y)))^2} \quad (3.26)$$

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. La figure 3.11 montre l'évolution des taux d'égalité erreur TEE en fonction du seuil Th par un filtre de *Gabor* avec résolution $\lambda=4$ et orientations $= \frac{\pi}{2}$.

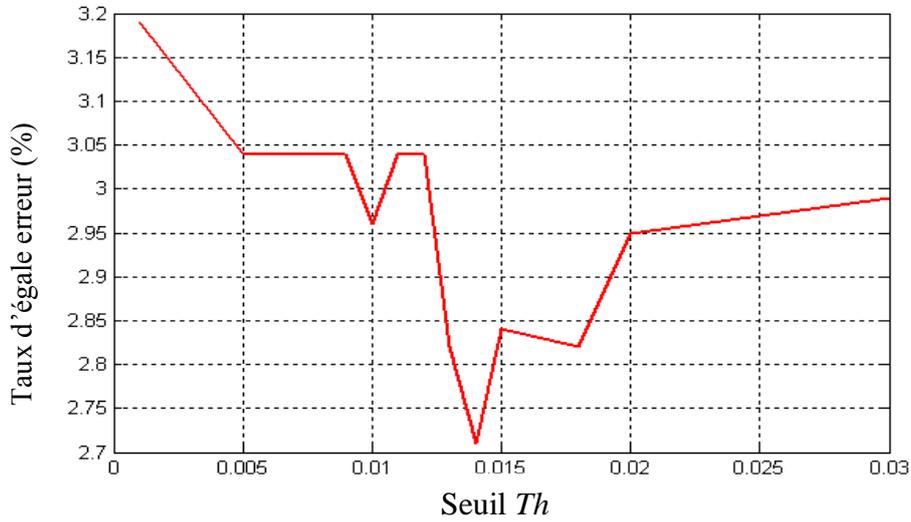


Figure 3.11: Taux d'égalité erreur dans l'ensemble d'évaluation en fonction du seuil Th .

La courbe de la figure 3.11 montre que la variation du taux de TEE en utilisant la phase de *Gabor* est bien liée au seuil du filtrage. Notre choix s'est porté sur le seuil de filtrage $Th = 0.014$ qui comme le montre la courbe présente le meilleur TEE. Notre seuil est presque égal à celui obtenu par *Mellakh* [257]. La même étude a été publiée dans [271] mais sur la base FRGC.

Dans cette deuxième phase, nous choisissons les filtres de *Gabor* optimaux pour la phase. Le tableau 3.4 présente les résultats de l'authentification de visages en terme de taux d'égalité erreur TEE dans l'ensemble d'évaluation pour différents niveaux de résolutions et d'orientations des quarante filtres de *Gabor* et un nombre de caractéristiques dans la phase de réduction d'espace par EFM, $N_p=100$.

Nous remarquons que la première résolution et les trois orientations: $\theta = \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{8}$ donnent le meilleur TEE. Les résultats obtenus par les phases de *Gabor* sont satisfaisants et encourageants. Nous utiliserons donc dans ce qui suit et pour la conception de notre système multi-algorithmes ces trois phases des filtres de *Gabor*. Les meilleurs systèmes d'authentification de visages sont présentés dans le tableau 3.5.

Tableau 3.4: Taux d'égale erreur dans l'ensemble d'évaluation pour différentes résolutions et orientations des phases pour les 40 filtres.

Résolution (λ)	Orientations des filtres (θ)							
	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$	$5\pi/8$	$3\pi/4$	$7\pi/8$
4	4.79	5.14	4.12	4.96	2.69	3.21	3.79	4.64
$4\sqrt{2}$	4.8	5.29	6	5.28	4.15	4.88	4.87	5.3
8	6.03	6.53	7.16	6.79	6.04	6.85	6.29	7.04
$8\sqrt{2}$	6.64	7.47	7.29	8.21	8.52	8.14	7.66	7.8
16	6.5	7.01	8.16	8.45	9.01	8.61	7.99	7.84

Tableau 3.5: Résultats du système d'authentification de visages pour les quatre meilleurs systèmes retenus dans l'ensemble d'évaluation et de test.

Méthodes	Ensemble évaluation	Ensemble de test		
	TEE %	TFR %	TFA %	TV %
Système 1	2.66	2	3.66	94.33
Système 2	2.69	0.5	4.07	95.43
Système 3	3.3	2	4.41	93.59
Système 4	3.79	0.5	4.47	95.03

Avec :

- Le **système 1** utilise l'étape de prétraitement (voir paragraphe 3.2.1) puis ACP + EFM comme étape de réduction d'espace et une comparaison par la métrique cosinus.
- Le **système 2** utilise la phase filtrée de la convolution de la composante S de l'espace de couleur TSV de l'image découpée, décimée et filtrée par le filtre de *Gabor* de la première résolution ($\lambda=4$) et l'orientation ($\theta = \pi/2$) puis ACP + EFM comme étape de réduction d'espace et enfin le cosinus pour la comparaison.
- Le **système 3** est identique au système 2 avec ($\lambda=4$) et l'orientation ($\theta = 5\pi/8$).
- Le **système 4** est aussi identique aux systèmes 2 et 3 avec ($\lambda=4$) et l'orientation ($\theta = 3\pi/4$).

Malgré des résultats satisfaisants pour les systèmes d'authentification de visages obtenus dans cette partie, les taux d'erreur associés à des systèmes biométriques uni-modaux restent relativement élevés. Ce qui les rend inacceptables pour un déploiement d'applications critiques de sécurité. Pour pallier à ces inconvénients, une solution est l'utilisation de plusieurs modalités biométriques au sein d'un même système. On parle alors de système biométrique multimodal qui est étudié dans le reste de ce chapitre.

3.3 Vérification multi-algorithmes de visage 2D

Comme il a été annoncé dans le résultat obtenu dans la section 3.2.5 concernant les systèmes uni-modaux d'authentification de visages, chaque système en soi ne peut pas toujours être utilisé de manière fiable pour effectuer la reconnaissance. Cependant, la consolidation d'informations présentées par les différentes modalités peut permettre une authentification précise de l'identité. Alors la biométrie multimodale est la combinaison de plusieurs modalités biométriques différentes, ce qui implique l'utilisation d'informations complémentaires pour une personne donnée. Dans cette section, nous allons présenter:

- les limitations des systèmes biométriques lorsqu'ils n'utilisent qu'une seule modalité,
- les différentes multi modalités possibles,
- l'état de l'art de la fusion de scores,
- la fusion de scores,

- les résultats.

3.3.1 Limitations des systèmes biométriques monomodaux

Bien que les techniques de reconnaissance biométrique promettent d'être très performantes, on ne peut garantir actuellement un excellent taux de reconnaissance avec des *systèmes biométriques unimodaux*. De plus, ces systèmes sont souvent affectés par les problèmes suivants [272] :

- **Bruit introduit par le capteur:** le bruit peut être présent dans les données biométriques acquises, ceci étant principalement dû à un capteur défaillant ou mal entretenu. Par exemple, un mauvais focus de caméra entraînant du flou dans des images de visage. Le taux de reconnaissance d'un système biométrique est très sensible à la qualité de l'échantillon biométrique et des données bruitées peuvent sérieusement compromettre la précision du système [273].
- **Non-universalité:** si chaque individu d'une population ciblée est capable de présenter une modalité biométrique pour un système donné, alors cette modalité est dite *universelle*. Ce principe d'universalité constitue une des conditions nécessaires de base pour un modèle de reconnaissance biométrique. Cependant, toutes les modalités biométriques ne sont pas vraiment universelles. Le *National Institute of Standards and Technologies (NIST)* a rapporté qu'il n'était pas possible d'obtenir une bonne qualité d'empreinte digitale pour environ **2%** de la population [274].
- **Manque d'individualité:** les caractéristiques extraites à partir de données biométriques d'individus différents peuvent être relativement similaires. Par exemple, une certaine partie de la population peut avoir une apparence faciale pratiquement identique à cause des facteurs génétiques (père et fils, vrais jumeaux, etc...). Ce manque d'unicité augmente le taux de fausse acceptation (TFA) d'un système biométrique.
- **Manque de représentation invariante:** les données biométriques acquises à partir d'un utilisateur lors de la phase de reconnaissance ne sont pas identiques aux données qui ont été utilisées pour générer le modèle de ce même utilisateur lors de la phase d'enrôlement. Ceci est connu sous le nom de *variations intra-classe*. Ces variations peuvent être dues à une mauvaise interaction de l'utilisateur avec le capteur (par exemple, changements de pose et d'expression faciale lorsque l'utilisateur se tient devant une caméra), à l'utilisation de capteurs différents lors de l'enrôlement et de la vérification, à des changements de conditions de l'environnement ambiant (par exemple, changements en éclairage pour un système de reconnaissance faciale). Cependant, dans la plupart des systèmes biométriques, ces caractéristiques ne sont pas invariantes et, par conséquent, des algorithmes complexes sont requis pour prendre en compte ces variations. De grandes *variations intra-classe* augmentent généralement le taux de faux rejet (TFR) d'un système biométrique.
- **Sensibilité aux attaques:** bien qu'il semble très difficile de voler les modalités biométriques d'une personne, il est toujours possible de contourner un système biométrique en utilisant des modalités biométriques usurpées. Des études [275] ont montré qu'il était possible de fabriquer de fausses empreintes digitales en gomme et de les utiliser pour contrer un système biométrique. Les modalités biométriques comportementales telles que la signature et la voix sont plus sensibles à ce genre d'attaque que les modalités biométriques physiologiques [78].

3.3.2 Les différentes multimodalités possibles

Les systèmes biométriques multimodaux diminuent les contraintes des systèmes biométriques monomodaux en combinant plusieurs systèmes. On peut différencier 5 types de systèmes multimodaux selon les systèmes qu'ils combinent (figure 3.12). On les appelle [276]:

- **multi-algorithmes:** lorsque plusieurs algorithmes traitent la même image acquise, cette multiplicité des algorithmes peut intervenir dans le module d'extraction en considérant plusieurs

ensembles de caractéristiques et/ou dans le module de comparaison en utilisant plusieurs algorithmes de comparaison.

- **multi-instances:** lorsqu'ils associent plusieurs instances de la même biométrie, par exemple l'acquisition de plusieurs images de visage avec des changements de pose, d'expression ou d'illumination.
- **multi-capteurs:** lorsqu'ils associent plusieurs capteurs pour acquérir la même modalité, par exemple un capteur optique et un capteur capacitif pour l'acquisition de l'empreinte digitale.
- **multi-échantillons:** lorsqu'ils associent plusieurs échantillons différents de la même modalité, par exemple deux empreintes digitales de doigts différents ou les deux iris. Dans ce cas les données sont traitées par le même algorithme mais nécessitent des références différentes à l'enregistrement contrairement aux systèmes multi-instances qui ne nécessitent qu'une seule référence.
- **multi-biométries:** lorsque l'on considère plusieurs biométries différentes, par exemple visage et iris.

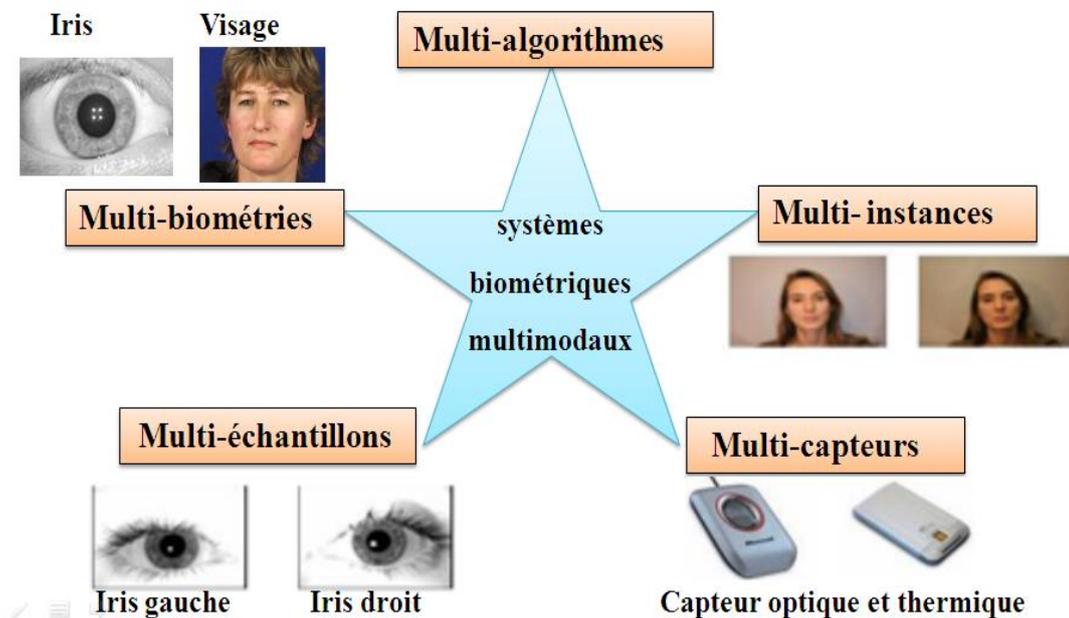


Figure 3.12: Les différents systèmes multimodaux.

La combinaison de plusieurs systèmes biométriques peut se faire à quatre niveaux différents: au niveau des **données**, au niveau des **caractéristiques** extraites, au niveau des **scores** issus du module de comparaison ou au niveau des **décisions** du module de décision.

Dans ce chapitre nous avons étudié la fusion au niveau de scores pour les raisons suivantes:

- La relation entre les espaces de caractéristiques des différents systèmes biométriques n'est pas forcément connue. Dans le cas où la relation est connue, on doit prendre soin d'éliminer les caractéristiques qui sont fortement corrélées. Cela requiert l'application d'algorithmes de sélection de caractéristiques avant l'étape de classification,
- La concaténation de deux vecteurs de caractéristiques peut engendrer un vecteur de caractéristiques ayant une grande dimension, menant au fameux problème de la "*malédiction de la dimensionnalité*". Bien que ce soit un problème général dans la plupart des applications de reconnaissance de forme, cela est encore plus marquant dans les applications biométriques à cause du temps, de l'effort et du coût impliqués dans la collecte de grandes quantités de données biométriques,
- La plupart des systèmes biométriques commerciaux ne fournissent pas l'accès aux vecteurs de caractéristiques qui sont utilisés dans leurs produits. Ainsi, très peu de chercheurs ont étudié la

fusion aux niveaux caractéristiques et la plupart d'entre eux se tournent généralement vers la fusion de scores,

- La fusion au niveau des scores est le type de fusion le plus utilisé car elle peut être appliquée à tous les types de systèmes (contrairement à la fusion pré-classification).
- Après les vecteurs de caractéristiques, les scores donnés en sortie par les systèmes contiennent l'information la plus riche à propos du modèle d'entrée.
- La fusion au niveau score donne le meilleur compromis entre la richesse d'information et la facilité d'implémentation. Aussi, est-il relativement facile d'accéder et de combiner les scores générés par les différents systèmes.

3.3.3 État de l'art de la fusion de scores

Au cours des dix dernières années les systèmes multi-algorithmes ou multi-classifieurs (multiple classifier systems *MCS*) sont devenus une méthode établie pour concevoir des systèmes de classification. Un grand nombre d'éléments de preuves à la fois théoriques et expérimentales montre que les systèmes multi-classifieurs peuvent surpasser un classifieur unique dans plusieurs applications réelles, en termes de précision de la classification (voir, par exemple [77, 36]). En particulier, plusieurs auteurs ont montré que *MCS* peut permettre d'améliorer l'authentification biométrique de visages [278, 279, 280].

Il existe deux approches pour fusionner les scores obtenus par différents classifieurs:

3.3.3.1 Approche par classification de scores

Plusieurs classifieurs ont été utilisés pour combiner les scores de correspondance afin d'arriver à une décision. *Wang et al* [281] considèrent les scores provenant de modules de reconnaissance faciale et de reconnaissance par l'iris comme un vecteur de caractéristiques à deux dimensions. Une analyse discriminante linéaire de *Fisher* (LDA) et un classifieur par réseau de neurones combiné à une fonction de base radiale (RBF) sont ensuite utilisés pour la classification. *Verlinde et Chollet* [282] combinent les scores provenant de deux modules de reconnaissance faciale et d'un module de reconnaissance de la parole avec l'aide de trois classifieurs. Le premier classifieur utilise la méthode des "k plus proches voisins" ("k nearest neighbor algorithm", "k-NN") avec une quantification vectorielle. Le deuxième classifieur est basé sur un arbre décisionnel et le dernier classifieur est basé sur un modèle de régression logistique.

Chatzis et al [283] utilisent une méthode de regroupement ("clustering") appelée "fuzzy k means" et une quantification vectorielle floue ("fuzzy vector quantization"), couplée à un classifieur de réseau de neurones de RBF médiane pour fusionner les scores obtenus à partir de systèmes biométriques basés sur des caractéristiques visuelles (visage) et acoustiques (voix). *Sanderson et al* [284] utilisent un classifieur basé sur une Machine à Vecteurs de Support (SVM) pour combiner les scores issus d'un module de reconnaissance faciale et d'un module de reconnaissance de la parole. Ils montrent que la performance d'un tel classifieur se détériore en la présence de conditions de bruit en entrée. Afin de surmonter ce problème, ils implémentent des classifieurs résistants au bruit structurel comme un classifieur linéaire défini par morceau ("Piece-Wise Linear Classifier") et un Classifieur Bayésien Modifié. *Ross et Jain* [285] utilisent un arbre décisionnel et des classifieurs discriminants linéaires pour combiner les scores des modalités du visage, d'empreinte digitale et de géométrie de la main.

3.3.3.2 Approche par combinaison de scores

Kittler et al [286] ont développé un cadre théorique pour combiner les informations d'identification obtenues à partir de plusieurs classifieurs en utilisant des schémas tels que la règle "somme", la règle "produit", la règle "maximum", la règle "minimum" et la règle "médiane". Afin d'employer ces schémas, les scores doivent être convertis en probabilités à postériori conformément à

un utilisateur client ou imposteur. Ils considèrent le problème de classer un modèle d'entrée X dans l'une des m classes possibles (dans un système de vérification, $m = 2$) d'après l'information d'identification fournie par R classifieurs différents.

Kwak et al [287] réalisent l'intégration floue de Choquet pour deux systèmes. Le premier système combine des images tirées du visage (yeux, nez et la bouche) et images globales du visage. Le deuxième système est basé sur des classifieurs opérant sur quatre ensembles de sous-images générées par décomposition en ondelettes. *Chia et al* [288] utilisent une méthode hybride de calcul du minimum, maximum ou somme pour la combinaison des scores de deux systèmes d'authentification de visages et voix. *Li et al* [289] utilisent la somme pondérée pour la fusion par combinaison des scores. *Shukla et al* [290] proposent un calcul adaptatif et hybride utilisant la combinaison de scores par logique floue à base de l'intégrale de *Sugeno* ou de *Choquet*.

Plus récents *Morizet* et *Allano* ont essayé d'utiliser les deux approches et arrivent à des taux de réussite très intéressants.

Allano [276] utilise les deux approches basées sur le SVM pour l'approche par classification de scores et la moyenne, le produit, le minimum, le maximum, la médiane et la somme pondérée pour l'approche par combinaison de scores. *Morizet* [78] utilise l'approche basée sur la combinaison de scores avec les méthodes simples (la moyenne, le produit, le minimum, le maximum, la médiane et la somme pondérée) et une nouvelle technique de fusion adaptative appelée "*Wavelet Score Denoising Statistical Fusion*" (WSDSF),

Nous nous sommes inspirés de leurs travaux et nous utilisons les deux approches dans l'objectif est de réduire le coût et le temps d'utilisation des systèmes multimodaux et améliorer les performances des systèmes biométriques.

3.3.4 La fusion de scores

Il existe deux approches pour combiner les scores obtenus par différents systèmes: la première est de traiter le sujet comme un problème de combinaison, tandis que la deuxième est de voir cela comme un problème de classification. Dans l'approche par combinaison, les scores individuels sont combinés de manière à former un unique score qui est ensuite utilisé pour prendre la décision finale. Afin de s'assurer que la combinaison de scores provenant de différents systèmes est cohérente, les scores doivent d'abord être transformés dans un domaine commun: on parle alors de *normalisation de score* [291].

3.3.4.1 Normalisation de scores

Les méthodes de normalisation de scores ont pour objectif de transformer individuellement chacun des scores issus des systèmes pour les rendre homogènes avant de les combiner. En effet, les scores issus de chaque système peuvent être de nature différente. Certains systèmes produisent des scores de similarité (plus le score est grand, plus la référence ressemble au test, donc l'utilisateur est un client), d'autres produisent des distances (plus la distance est faible, plus la référence et le test sont proches, plus l'utilisateur est un client). De plus chaque système peut avoir des intervalles de variations des scores différents. Par exemple, pour un système, les scores varient entre 0 et 1 et pour un autre les scores varient entre 0 et 1000. On comprend bien la nécessité de normaliser les scores avant de les combiner. Les méthodes de normalisation, présentées dans la suite, traitent des scores qui varient tous dans le même sens (en général on considère tous les scores sous forme de similarité [76]).

Les différentes techniques de normalisation de scores utilisées sont :

- **Normalisation par la méthode Min-Max [76, 78, 292]**

Soit s_{ij} le $j^{ème}$ score de sortie du $i^{ème}$ système, où $i = 1, 2, \dots, N$ et $j = 1, 2, \dots, M$ (N est le nombre de systèmes et M le nombre de scores disponibles dans l'ensemble de données d'entraînement). Le score normalisé Min-Max pour le score de test s_{ik} est donné par :

$$s'_{ik} = \frac{s_{ik} - \min(\{s_i\})}{\max(\{s_i\}) - \min(\{s_i\})} \quad (3.27)$$

Où $\{s_i\} = \{s_{i1}, s_{i2}, \dots, s_{iM}\}$. La normalisation Min-Max conserve la distribution de scores originale à un facteur d'échelle près et transforme tous les scores dans l'intervalle $[0, 1]$.

- **Normalisation par la méthode Z-Score [276, 78, 292]**

La technique de normalisation de score la plus employée est certainement la Z-Score qui utilise la moyenne arithmétique (μ) et l'écart-type (σ) des données. On peut s'attendre à ce que cette méthode fonctionne bien si on a une connaissance à priori du score moyen et des variations de scores d'un système. Les scores normalisés sont donnés par:

$$s'_{ik} = \frac{s_{ik} - \mu}{\sigma} \quad (3.28)$$

- **Normalisation par la méthode tangente hyperbolique "Tanh" [276, 78, 293]**

Les scores normalisés sont donnés par (3.29):

$$s'_{ik} = \frac{1}{2} \left(\tanh \left(0.001 \frac{s_{ik} - \mu}{\sigma} \right) + 1 \right) \quad (3.29)$$

Où μ est la moyenne arithmétique, σ l'écart-type des données et \tanh la tangente hyperbolique. La méthode tangente hyperbolique met chaque score normalisé dans l'intervalle $[0, 1]$.

- **Normalisation par la médiane et l'écart absolu médian (MAD) [78, 292]**

Ces deux méthodes sont insensibles aux valeurs aberrantes et aux points aux extrémités d'une distribution. Ainsi, une méthode de normalisation utilisant la médiane et la MAD (l'écart absolu médian) serait robuste et est donnée par :

$$s'_{ik} = \frac{s_{ik} - \text{median}}{MAD}, \quad MAD = \text{median}(|s_i - \text{median}(s_i)|) \quad (3.30)$$

- **Normalisation par une fonction Quadratique-Linéaire-Quadratique (QLQ) [78, 294, 295, 296]**

Snelick et al [294] utilisent une fonction Quadratique-Linéaire-Quadratique (QLQ) pour normaliser des scores au préalable transformés dans l'intervalle $[0, 1]$ grâce à une normalisation Min-Max (s_{MM}). Cette normalisation QLQ prend comme paramètres le centre c et la largeur w de la zone de recouvrement des distributions des scores Imposteurs et Client (figure 3.13).

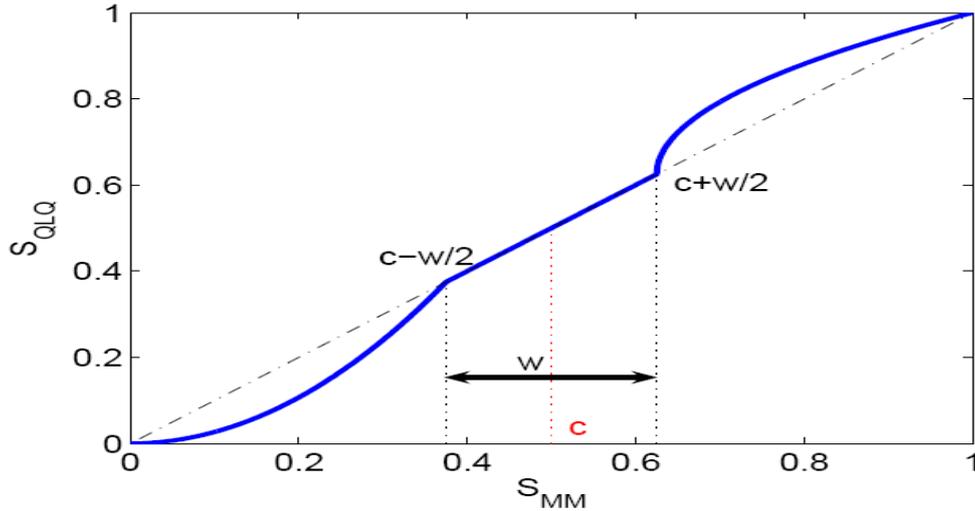


Figure 3.13: Normalisation QLQ.

La zone de recouvrement reste inchangée tandis que les autres régions sont transformées à l'aide de deux fonctions quadratiques par segment. Le score normalisé est donné par:

$$s_{QLQ} \begin{cases} \frac{1}{c-\frac{\omega}{2}} s_{MM}^2, & si \quad s_{MM} \leq \left(c - \frac{\omega}{2}\right) \\ s_{MM}, & si \quad \left(c - \frac{\omega}{2}\right) < s_{MM} \leq \left(c + \frac{\omega}{2}\right) \\ \left(c + \frac{\omega}{2}\right) + \sqrt{\left(1 - c - \frac{\omega}{2}\right) \left(s_{MM} - c - \frac{\omega}{2}\right)}, & si \quad si \quad non \end{cases} \quad (3.31)$$

- **Normalisation par une fonction double sigmoïde [78, 296]**

Cappelli et al [295] ont utilisé une fonction double sigmoïde pour la normalisation de scores dans un système biométrique multimodal qui combine différents systèmes d'empreinte digitale. Le score normalisé est donné par:

$$s'_{ik} = \begin{cases} \frac{1}{1 + \exp\left(-2\left(\frac{s_{ik}-t}{r_1}\right)\right)} & si \quad s_k < t \\ \frac{1}{1 + \exp\left(-2\left(\frac{s_{ik}-t}{r_2}\right)\right)} & si \quad non \end{cases} \quad (3.32)$$

Où t est le point de fonctionnement de référence et r_1 et r_2 sont respectivement les bords gauche et droit.

Ce schéma transforme les scores dans l'intervalle $[0,1]$. Mais il demande de paramétrer prudemment t , r_1 et r_2 afin d'obtenir une bonne efficacité. En général, t est choisi de façon à être une valeur appartenant à la région de recouvrement entre les distributions des scores client et imposteurs. r_1 et r_2 sont fixés de telle sorte qu'ils correspondent respectivement aux extrémités de recouvrement entre les deux distributions vers la gauche et vers la droite de t (on peut également voir cela comme les valeurs correspondant respectivement au minimum des scores de similarité des clients et au maximum des scores de similarité des imposteurs). Ce schéma de normalisation fournit donc une transformation linéaire des scores dans la région de recouvrement, tandis que les scores à l'extérieur de cette région sont transformés de manière non linéaire.

3.3.4.2 Approche par combinaison de scores

- **Méthode de combinaisons simples**

Les méthodes de combinaisons de scores simples sont des méthodes très simples dont l'objectif est d'obtenir un score final s à partir des N scores disponibles s_i pour $i = 1$ à N issus de N systèmes. Les méthodes utilisées sont la moyenne, le produit, le minimum, le maximum, la médiane et la somme pondérée présentés respectivement dans [276, 289].

$$s = \frac{1}{N} \sum_{i=1}^N s_i \tag{3.33}$$

$$s = \prod_{i=1}^N s_i \tag{3.34}$$

$$s = \min(s_i) \tag{3.35}$$

$$s = \max(s_i) \tag{3.36}$$

$$s = \text{median}(s_i) \tag{3.37}$$

$$s = \sum_{i=1}^N w_i s_i \tag{3.38}$$

La somme pondérée permet de donner des poids différents w_i à chacun des systèmes en fonction de leur performance individuelle ou de leur intérêt dans le système multi-algorithmes.

- **Combinaison de scores par logique floue**

La théorie de la logique floue (des sous-ensembles flous) a été introduite par Zadeh en 1965 [297] comme une extension de la logique binaire d'une part et une amélioration de la logique multivaluée (multi-valued logics en anglais) d'autre part. L'importance de la logique floue réside dans le fait qu'elle s'approche du raisonnement humain par l'intégration et le traitement du caractère approximatif, vague, imprécis ou flou de la connaissance humaine. Les termes linguistiques tels que «environ », « moyenne », « approximativement » sont de nature à donner un caractère flou aux phrases énoncées. Par exemple, la règle « si le prix est inférieur à 6000 DA, j'achète » sera intuitivement utilisable si le prix est de 6002 DA, mais elle ne pourrait être exploitée en logique classique puisque le prix indiqué ne satisferait pas la prémisse.

- *Mesure floue*

Un jeu de fonctions $g: P(Y) \rightarrow [0, 1]$ est appelée une mesure floue si les conditions suivantes sont remplies :

- 1- conditions aux limites: $g(\emptyset) = 0, g(Y) = 1$
- 2- monotonie : $g(A) \leq g(B), si A \subset B et A, B \in P(Y)$
- 3- continuité : $\lim_{i \rightarrow \infty} g(A_i) = g\left(\lim_{i \rightarrow \infty} A_i\right), si \{A_i\}_i^\infty$ est une suite croissante d'ensembles mesurables.

A partir de cette définition, Sugeno [298] a introduit un soi-disant g_λ mesure floue qui est livré avec une propriété supplémentaire

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B) \tag{3.39}$$

Pour tous les $A, B \subset Y$ et $A \cap B = \emptyset$, et pour certains $\lambda > -1$.

Évidemment quand $\lambda=0$, le g_λ mesure floue devient une mesure de probabilité standard.

En général, la valeur de λ peut être déterminée en raison de l'état limite de la mesure floue g_λ . Cette condition pour $(Y) = 1$. Par conséquent, la valeur de λ est déterminée par la résolution de ce qui suit:

$$g_\lambda(Y) = \frac{1}{\lambda} (\prod_{i=1}^n (1 + \lambda g^i) - 1), \quad \prod_{i=1}^n (1 + \lambda g^i) - 1 \neq 0 \quad (3.40)$$

est l'équivalent de :

$$\lambda + 1 = \prod_{i=1}^n (1 + \lambda g^i) \quad (3.41)$$

Où $\lambda \in (-1, +\infty)$, $\lambda \neq 0$, et g^i est la valeur de la fonction de densité floue. La solution peut être facilement obtenue; on s'intéresse évidemment à la racine supérieure à -1 [299].

- *Intégrale floue*

L'intégrale floue de la fonction h calculée sur Y par rapport à une mesure floue g est définie sous la forme:

$$\int_Y h(y) o g(\cdot) = \sup_{\alpha \in [0,1]} [\min[\alpha, g(\{y | h(y) \geq \alpha\})]] \quad (3.42)$$

Intégrale floue de Sugeno [287, 290, 300]

Lorsque les valeurs des $h(\cdot)$ (les scores) sont classés dans l'ordre décroissant, $h(y_1) \geq h(y_2) \geq \dots \geq h(y_n)$

L'intégrale floue de Sugeno est calculée comme suit:

$$\int_Y h(y) o g(\cdot) = \max_{i=1:n} [\min(h(y_i), g(A_i))] \quad (3.43)$$

Où $A_i = \{y_1, y_2, \dots, y_i\}$ désigne un sous-ensemble d'éléments. Les valeurs de $g(A_i)$ pris en charge par la mesure floue sur les sous-ensembles correspondant d'éléments peut être déterminées de manière récursive sous la forme :

$$g(A_i) = g(y_i) = g^1 \quad (3.44)$$

$$g(A_i) = g^i + g(A_{i-1}) + \lambda g^i g(A_{i-1}) \quad (3.45)$$

Le calcul de la fonction de densité floue g^i sur la base des données est assuré par la manière suivante:

$$\begin{cases} g^i = \beta p_i & i = 1 \\ g^i = (1 - \beta) p_i, & i = 2,3,4 \end{cases} \quad (3.46)$$

Où p_i est le taux de classification dans l'intervalle $[0, 1]$ pour chaque système.

$B \in [0, 1]$ est un facteur qui met en place un certain équilibre entre les résultats de la classification.

Intégrante floue de Choquet [287, 290, 300]

Il a été démontré que (3.43) n'est pas une extension correcte de l'intégration de Lebesgue habituelle. En d'autres termes, lorsque la mesure est additive l'expression ci-dessus ne retourne pas

l'intégrale au sens de *Lebesgue*. Afin de remédier à cet inconvénient, *Murofushi* et *Sugeno* [301] ont proposés une soi-disant intégrante floue de *Choquet* calculé de la manière suivante:

$$\int_Y h(y)dg(\cdot) = \sum_{i=1}^n [h(y_i) - h(y_{i+1})]g(A_i)h(y_{n+1}) = 0 \quad (3.47)$$

3.3.4.3 Approche par classification de scores

L'approche par classification de scores [302] cherche à prendre une décision à partir du vecteur de scores de dimension N . Le but des méthodes de fusion basées sur des classifieurs est de chercher une séparatrice entre les 2 classes, Client et Imposteur, dans l'espace à N dimensions des scores. La séparatrice peut être plus ou moins complexe selon le classifieur utilisé. Toutes les méthodes classiques de reconnaissance des formes et d'apprentissage automatique peuvent être utilisées pour la fusion de scores qui est un problème de classification comme un autre. Dans le domaine des systèmes multi-algorithmes, dont l'objectif est d'utiliser un classifieur pour combiner plusieurs systèmes sources (ici ce sont des systèmes biométriques que l'on veut fusionner) afin de consolider la décision de classification. Un grand nombre de méthodes ont été utilisées. Les méthodes utilisées dans notre travail sont :

- **Fusion par méthode statistique de *Fisher***

La méthode statistique introduite ici, se base sur les travaux de *Fischer* [96] et fait usage d'une frontière de décision linéaire pour séparer deux populations données, à savoir les clients et les imposteurs dans notre cas. Dans le contexte particulier de l'authentification de personnes, *Duc* et *al* [303] utilisent eux aussi une règle de décision linéaire, mais l'optimisent en se fixant un critère dérivé de la théorie de *Bayes* [304].

Envisageons à présent la règle de décision développée par *Fisher*. Elle se base sur le rapport de vraisemblance repris ci-dessous:

$$\frac{T(z|c)}{T(z|i)} > k \quad (3.48)$$

Où k représente un seuil d'acceptation dont la valeur dépend du compromis *TFA/TFR* que l'on veut obtenir. Dans le problème qui nous préoccupe, $T(z|c)$ et $T(z|i)$ sont inconnus et doivent être estimés à partir des données d'apprentissage. Une hypothèse courante consiste à approcher les distributions réelles par des distributions normales à p variables $N_p(\mu_A, \Sigma)$, où $A = \{c, i\}$ représente la classe d'individus, μ_A le vecteur des scores moyens et Σ la matrice de covariance entre experts. Dans un premier temps, on supposera la matrice Σ indépendante de la classe d'individus. Sous de telles hypothèses, les fonctions de densité de probabilité s'écrivent sous la forme:

$$f_A(z) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (z - \mu_A)' \Sigma^{-1} (z - \mu_A) \right\} \quad (3.49)$$

Les paramètres μ_c , μ_i et Σ sont inconnus, mais peuvent être estimés à partir des données d'apprentissage, soit x les μ_c données relatives aux accès clients et y , les n_i données relatives aux accès imposteurs (simulés). On a:

$$\hat{\mu}_c = \sum_{q=1}^{n_x} \frac{x_q}{n_c} \quad (3.50)$$

$$\hat{\mu}_i = \sum_{q=1}^{n_y} \frac{y_q}{n_i} \quad (3.51)$$

$$\widehat{\Sigma}_c = \sum_{q=1}^{n_x} (x_q - \widehat{\mu}_c)(x_q - \widehat{\mu}_c)' / (n_c - 1) \quad (3.52)$$

$$\widehat{\Sigma}_i = \sum_{q=1}^{n_y} (y_q - \widehat{\mu}_i)(y_q - \widehat{\mu}_i)' / (n_i - 1) \quad (3.53)$$

$$\widehat{\Sigma} = \frac{[(n_c-1)\widehat{\Sigma}_c + (n_i-1)\widehat{\Sigma}_i]}{n_c + n_i - 2} \quad (3.54)$$

Notons que l'on tient compte ici, par l'intermédiaire de Σ , de la dépendance qui peut exister entre experts. En combinant les équations (3.48) à (3.54), on peut réécrire $\widehat{f}_c(z)/\widehat{f}_i(z) \geq k$ sous la forme de $D_L(z) \geq \ln(k) = k^*$ où

$$D_L(z) = \left(z - \frac{1}{2}(\widehat{\mu}_c + \widehat{\mu}_i) \right)' \widehat{\Sigma}^{-1} (\widehat{\mu}_c - \widehat{\mu}_i) \quad (3.55)$$

Fisher fut le premier à utiliser cette fonction à des fins de classification. Comme $D_L(z)$ est linéaire en z , elle fut communément appelée fonction linéaire discriminante ou LDF (Linear Discriminant Function). Ainsi, la procédure à suivre pour vérifier l'identité d'un candidat, consiste à calculer $\widehat{\mu}_c$, $\widehat{\mu}_i$ et $\widehat{\Sigma}$ à partir des données d'entraînement (ce qui est effectué une fois pour toutes), puis $D_L(z)$ et comparer celui-ci au seuil k^* donné. Si $D_L(z) \geq k^*$, le candidat est accepté comme client.

Dans le cas de distributions de scores clients et imposteurs ne satisfaisant pas l'hypothèse d'une covariance Σ unique, la règle de décision (3.55) peut se réécrire sous la forme $D_Q \geq 2k^*$, où

$$D_Q(z) = (z - \widehat{\mu}_i)' \widehat{\Sigma}_i^{-1} (z - \widehat{\mu}_i) - (z - \widehat{\mu}_c)' \widehat{\Sigma}_c^{-1} (z - \widehat{\mu}_c) + \ln\left(\frac{|\widehat{\Sigma}_i|}{|\widehat{\Sigma}_c|}\right) \quad (3.56)$$

$D_Q(z)$ est appelée fonction quadratique discriminante ou QDF (Quadratic Discriminant Function). Cette règle permet d'obtenir de meilleurs résultats que ceux obtenus dans le cadre d'une fusion linéaire, mais requiert un ensemble d'apprentissage suffisamment étendu pour pouvoir évaluer Σ_c et Σ_i avec précision [305].

- Les machines à vecteurs de support (SVM)

Les machines à vecteurs de support (Support Vector Machine: SVM), également appelées Séparateurs à Vaste Marges, introduits par *Vapnik* [306] et utilisées dans la fusion de données dans les références [276, 296, 307, 308, 309, 310]. Ils ont pour but de définir un hyperplan de séparation entre les deux classes, hyperplan qui minimise l'erreur de classification sur un ensemble d'apprentissage. Cette méthode d'apprentissage supervisée peut apprendre une séparatrice plus ou moins complexe selon la nature du noyau choisi. Le noyau le plus simple est le noyau linéaire qui consiste à chercher une séparatrice linéaire dans l'espace à N dimensions des scores. Le but des fonctions noyau est de transformer l'espace initial (des scores à N dimensions) en un espace de dimension plus grande dans lequel les données pourraient être linéairement séparables. La séparatrice est donc toujours linéaire dans l'espace transformé par la fonction noyau, mais ne l'est plus dans l'espace des scores.

Le but de la SVM est donc de trouver une séparatrice qui minimise l'erreur de classification sur l'ensemble d'apprentissage mais qui sera également performante en général sur des données non utilisées en apprentissage. Pour cela, le concept utilisé est celui de marge (d'où le nom de séparateurs à vaste marge). La marge est la distance quadratique moyenne entre la séparatrice et les éléments d'apprentissage les plus proches de celle-ci appelés vecteurs de support (figure 3.14). Ces éléments sont appelés vecteurs de support car c'est uniquement sur ces éléments de l'ensemble d'apprentissage que la séparatrice est optimisée.

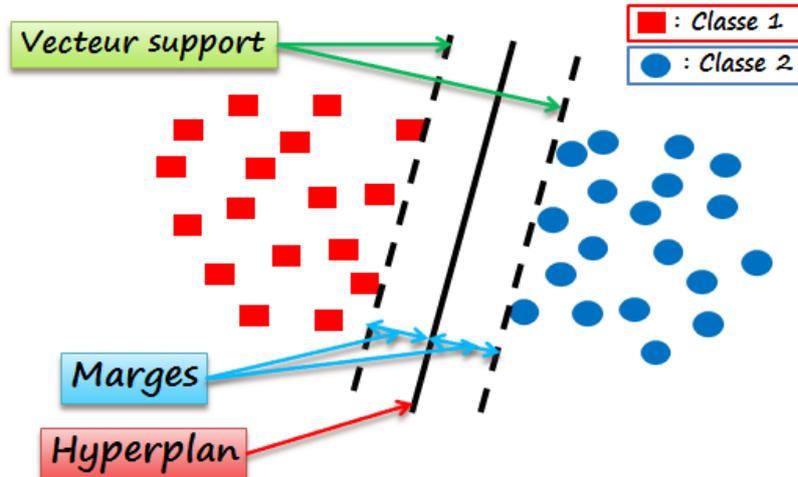


Figure 3.14: Séparation linéaire dans un espace à deux dimensions.

Tout classifieur a pour but de classer un élément x , ici $x = (s_1, \dots, s_N)$ est un vecteur de scores de dimension N dans l'une des classes possibles. Dans notre problème, il y a deux classes, Client ou Imposteur, dont l'étiquette sera noté y avec $y = -1$ et 1 . La valeur -1 correspondant à la classe des Imposteurs et 1 à la classe des Clients. Le classifieur a donc pour but de déterminer f telle que :

$$y = f(x) \tag{3.57}$$

La SVM a pour but de trouver la meilleure séparatrice linéaire (en terme de marge maximale, c'est à dire la meilleure généralisation) dans l'espace transformé par la fonction de noyau K . Cela consiste à déterminer le vecteur w et la constante b tels que la séparatrice ait pour équation:

$$wK(x) + b = 0 \tag{3.58}$$

La distance entre un point de l'espace x_i et l'hyperplan d'équation $wK(x) + b = 0$ est égale à:

$$h(x_i) = \frac{w \cdot K(x_i) + b}{\|w\|} \tag{3.59}$$

Pour maximiser la marge, il faut donc minimiser $\|w\|$ tout en maximisant $wK(x) + b$ pour les x_i définis comme vecteurs de support. Ces vecteurs de supports sont les x_i pour $i = 1 : m$ de la base d'apprentissage tels que $wK(x) + b = \pm 1$.

La résolution de ce problème d'optimisation est faite par l'utilisation des multiplicateurs de *Lagrange* où le Lagrangien est donné par :

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^m \alpha_i (y_i (w \cdot K(x_i) + b) - 1) \tag{3.60}$$

avec les coefficients α_i appelés multiplicateurs de *Lagrange*.

Pour résoudre ce problème d'optimisation, il faut minimiser le Lagrangien par rapport à w et b et le maximiser par rapport à α . Dans la pratique, il est souvent impossible de trouver une séparatrice linéaire (même dans l'espace transformé par la fonction noyau) car il y a toujours des erreurs de classification. *Vapnik* [306] a donc introduit la technique de "marge souple". Ce principe de marge souple tolère les mauvais classements par l'introduction de variables ressorts ζ_i qui permettent de relâcher les contraintes sur les éléments d'apprentissage qui ne doivent plus être à une distance

supérieure ou égale à 1 de la marge (l'égalité correspondant aux vecteurs de support), mais à une distance supérieure ou égale à $1 - \zeta_i$, c'est à dire:

$$y_i(w \cdot K(x_i) + b) \geq 1 - \zeta_i \quad (3.61)$$

Avec $\zeta_i \geq 0$ pour $i = 1$ à M , M étant le nombre d'éléments de la base d'apprentissage.

Le problème d'optimisation est donc modifié et le Lagrangien devient :

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - C \sum_{i=1}^M \zeta_i - \sum_{i=1}^M \alpha_i (y_i(w \cdot K(x_i) + b) - 1) \quad (3.62)$$

Où C est une constante strictement positive qui permet de régler le compromis entre le nombre d'erreurs de classification et la largeur de la marge [276].

- Réseaux de neurones artificiels

Le principe général des Réseaux de Neurones Artificiels (RNA) est à l'origine inspiré de certaines fonctions de base des neurones naturels du cerveau.

Un réseau de neurones artificiel est généralement organisé en plusieurs couches :

- une couche d'entrée,
- une couche de sortie,
- des couches intermédiaires appelées couches cachées.

La présence de couches cachées permet de discriminer des classes d'objets non linéairement séparables.

En général, un réseau de neurones est fondamentalement un classifieur. Il réalise un travail de classification pendant la phase d'apprentissage et de classement lors de la reconnaissance. Mais on peut l'utiliser pour réaliser la fusion de données [310, 311] pour séparer deux populations données, à savoir les clients et les imposteurs dans notre cas.

Les grands avantages des réseaux de neurones résident dans leur capacité d'apprentissage automatique (approximation universelle (*Cybenko, Hornik*)), ce qui permet de résoudre des problèmes sans nécessiter l'écriture de règles complexes, tout en étant tolérant aux erreurs. Ils résident aussi dans leur capacité à prendre une décision à partir de critères non formalisables explicitement [312].

3.3.5 Résultat

Le système multimodal étudié dans cette section est un système multi-algorithmes ou multi-classifieurs (Multiple Classifier Systems). Celui-ci est construit par les quatre systèmes retenus dans la section 3.2.5. En effet, lorsque l'on cherche à fusionner plusieurs systèmes d'authentification de visages, comment choisir la méthode la plus appropriée ? Le choix de cette méthode dépend de plusieurs critères qui sont :

- les systèmes à combiner et surtout les distributions de leurs scores,
- la difficulté des données, c'est-à-dire la séparabilité des classes Client et Imposteur,
- la quantité de données disponibles pour l'apprentissage,
- l'application visée.

Selon la nature des scores de comparaison issus de chaque système, une étape de normalisation des scores peut être nécessaire. Cette étape ne pose pas vraiment de problème à partir du moment où l'on connaît la loi des scores ou que l'on possède un échantillon d'apprentissage pour l'estimer.

La difficulté des données est associée à la possibilité de séparer les classes client et imposteur. Cette caractéristique est directement liée à la performance de reconnaissance. La difficulté des données doit être considérée séparément pour chaque système biométrique à combiner pour une connaissance préalable de l'étape de fusion des différentes performances individuelles des systèmes. En effet, on ne

combinera pas de la même façon des systèmes avec des performances à peu près équivalentes et des systèmes avec des performances très différentes.

Dans ce dernier cas, on voudra que le "bon" système ait plus de poids dans la décision que le "mauvais". Ainsi, l'étude des performances individuelles peut nous aider à choisir la méthode de fusion adéquate. Mais c'est surtout la séparabilité des données multidimensionnelles (c'est-à-dire en considérant les vecteurs de scores) qui va nous donner une indication sur la complexité de la méthode de fusion à utiliser. En effet, des données très bien séparées dans l'espace des scores à N dimensions ne nécessiteront pas l'utilisation d'une méthode très complexe, une séparatrice linéaire sera suffisante. En revanche, dans le cas de données non séparables linéairement, on pourra essayer d'utiliser une méthode de fusion plus complexe avec une séparatrice ou des transformations de scores non linéaires.

L'étude de la séparabilité des données ou l'utilisation de méthodes de fusion complexes nécessitent la disponibilité de données étiquetées pour l'apprentissage selon la quantité et la représentativité des données disponibles.

L'application visée a également une part importante dans le choix de la méthode de fusion. En revanche, les exigences de performance sont impactées par le choix de la méthode de fusion. En effet, selon les applications (de sécurité ou de confort), les performances attendues ne sont pas les mêmes et surtout les deux types d'erreurs de classification que sont le taux de fausse acceptation et le taux de faux rejet ne sont pas considérés de la même façon. Selon le point de fonctionnement visé, on ne choisira pas les mêmes méthodes de fusion. Par exemple, un classifieur comme l'optimise sa séparatrice en considérant de la même façon les deux types d'erreurs n'est donc peut être pas approprié pour une application de sécurité qui cherche à minimiser les fausses acceptations.

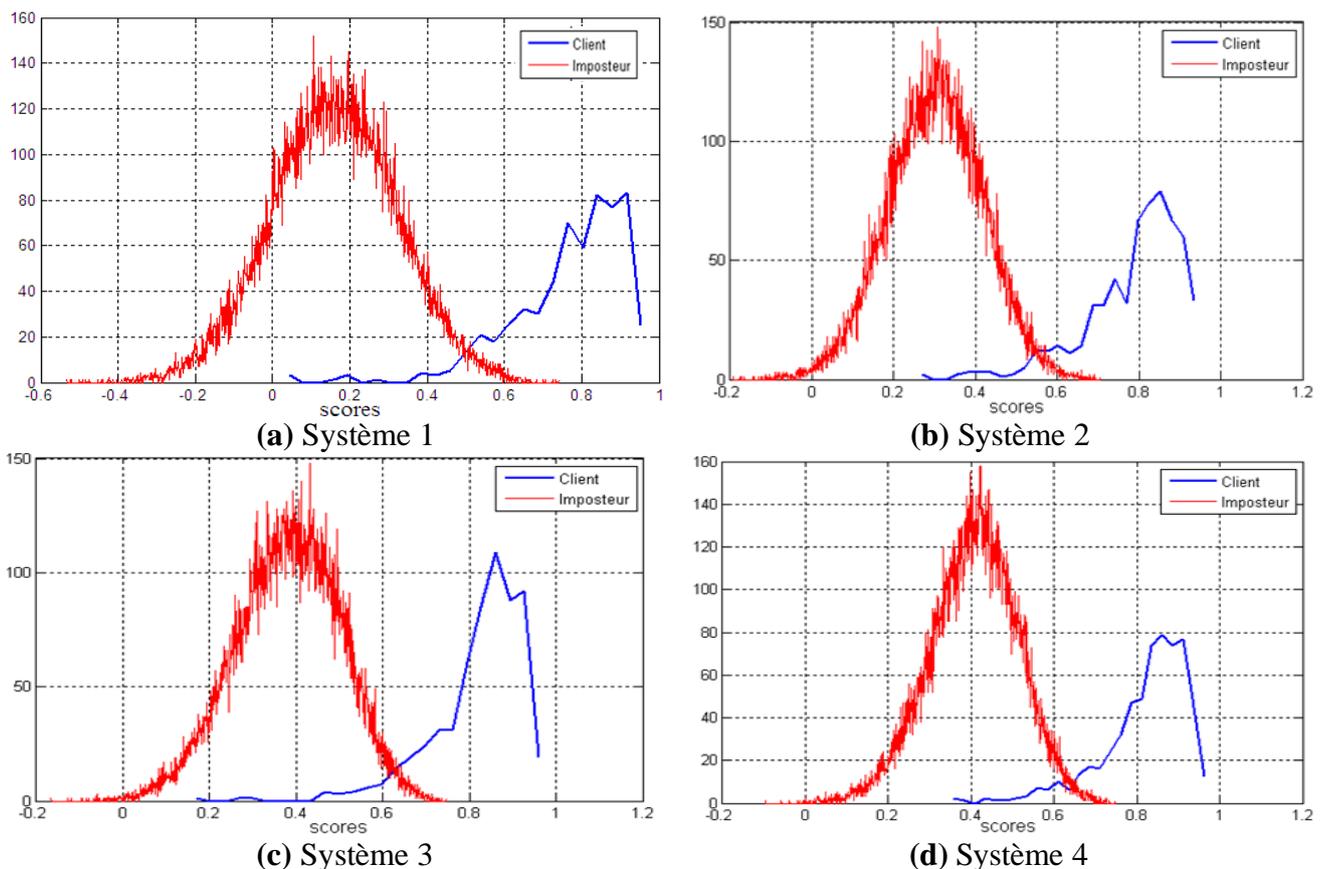


Figure 3.14: Distributions de scores des quatre systèmes d'authentification de visages sur la base de données XM2VTS.

Cette section a donc pour but de comparer toutes les méthodes de normalisation des scores et toutes les méthodes de fusion de scores pour les deux approches combinaison et classification présentées à la

section 3.3.4 pour les quatre systèmes d'authentification de visages appliqués sur la base de données XM2VTS.

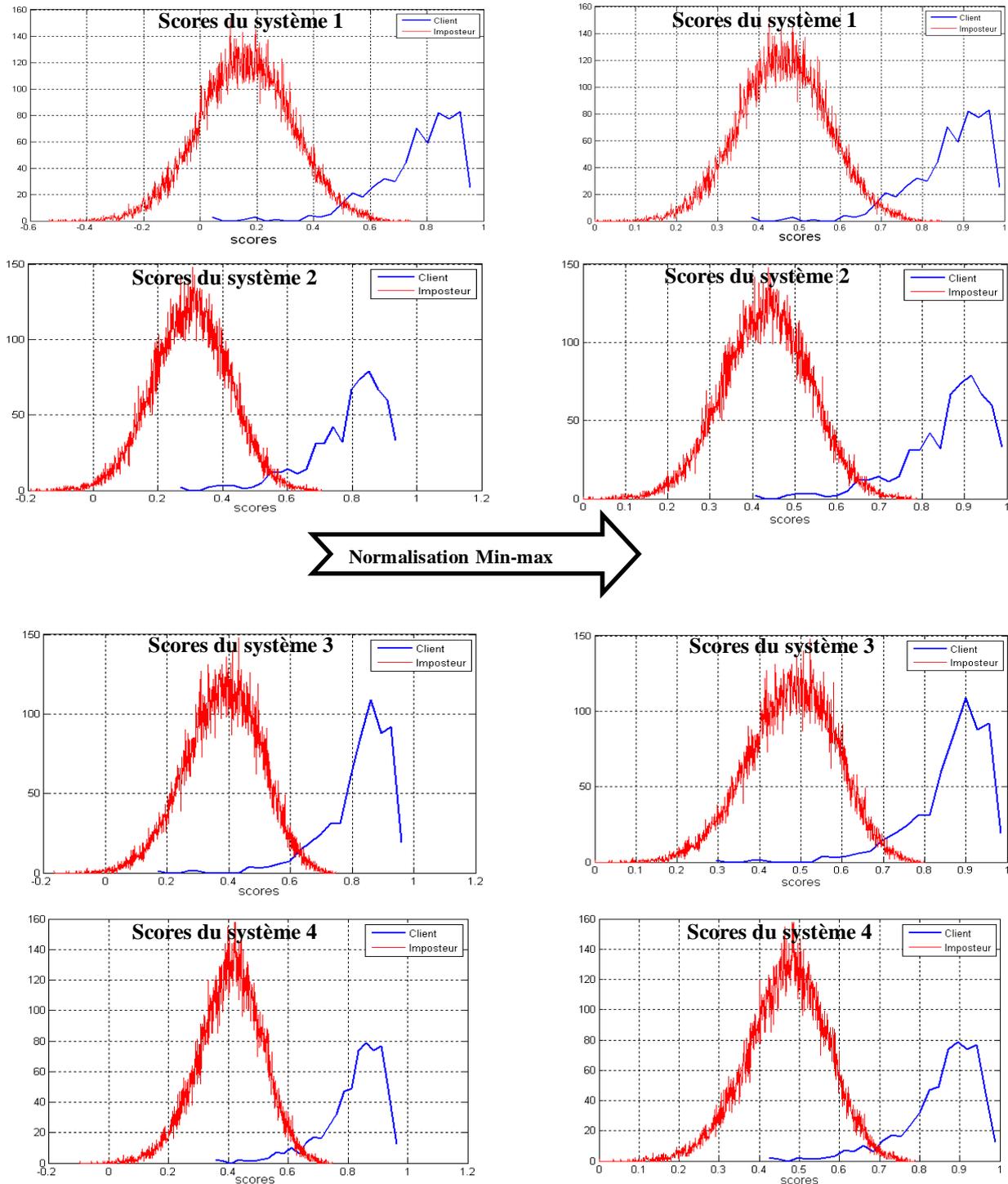


Figure 3.15: Normalisation de scores par la méthode du Min-Max.

3.3.5.1. Les distributions et normalisation de scores

Les distributions de scores pour les quatre systèmes d'authentification de visages sont représentées sur la figure 3.14. Nous remarquons que les quatre systèmes donnent des distributions Client et Imposteur différentes. Elles sont différentes en termes de plage de variation, ce qui rendra

Chapitre 3 : Vérifications uni-modale et multi-algorithmes de visage 2D

nécessaire l'étape de normalisation des scores détaillée à la section 3.3.4.1. Elles sont également différentes dans leurs formes et leurs recouvrements entre les deux classes (Client et Imposteur), même si toutes les distributions ont un seul mode (mais pas forcément symétriques).

On peut voir par exemple l'effet de la normalisation Min-Max pour les quatre systèmes sur la figure 3.15. Nous remarquons que toutes les distributions normalisées sont dans un intervalle conjoint et la variation des distributions se trouve entre 0 et 1.

Dans la figure 3.16 la transformation des scores du premier système est présentée pour les six normalisations.

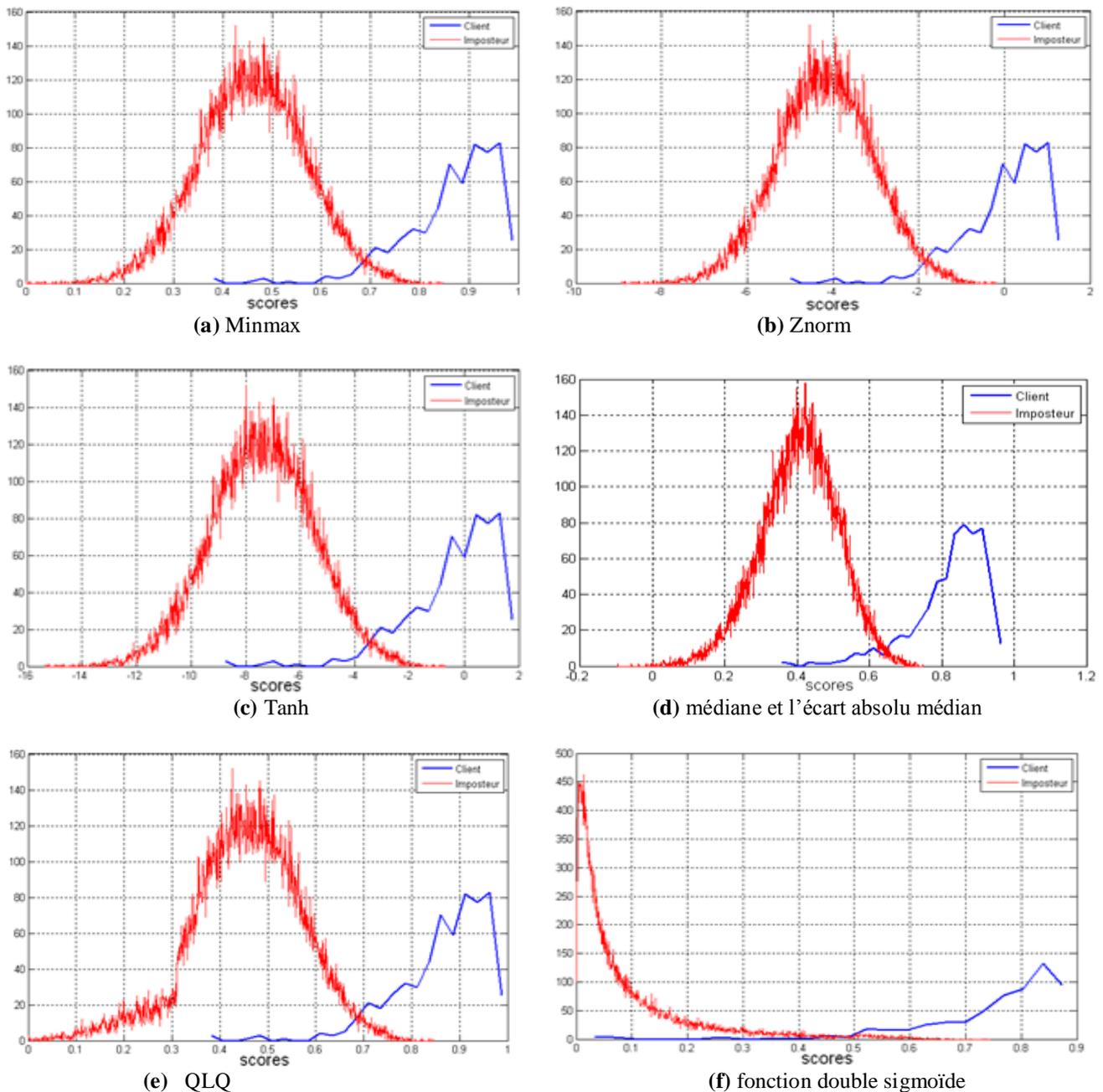


Figure 3.16: Normalisation de scores.

Nous remarquons que les méthodes de normalisation Min-Max, Znorm, Tanh, la médiane et l'écart absolu médian ne modifient pas la forme des distributions mais leur différence se situe dans la façon dont chaque distribution sera répartie dans l'intervalle et donc dans la façon dont les systèmes à combiner se juxtaposeront dans l'intervalle défini (avec Znorm la moyenne de distribution des Clients

de chaque système qui sera égale à 0). Alors qu'avec Min-Max ce sont les extrémités qui sont définies et la moyenne sera positionnée différemment selon les systèmes. De plus, les méthodes de normalisation par fonction Quadratique-Linéaire-Quadratique (QLQ) et la fonction double sigmoïde modifient la forme des distributions.

3.3.5.2 Comparaison des méthodes de normalisation associées aux méthodes de combinaisons

- Méthodes de normalisation associées aux méthodes de combinaisons simples

Nous avons vu dans le paragraphe précédent que les scores ont des distributions et intervalles de variation différents et qu'il est donc nécessaire de les normaliser avant de les fusionner. Comme nous l'avons déjà cité, il en existe deux types:

- des méthodes de normalisations de scores qui ne modifient pas la forme des distributions,
- les méthodes qui modifient la forme des distributions.

Ces méthodes de normalisation sont ensuite associées aux méthodes de combinaison de scores simples que sont la somme, le produit, le minimum, le maximum, la médiane et la somme pondérée. En effet, pour réaliser la fusion il faut combiner les quatre scores normalisés pour obtenir un score final qui permette de calculer les taux d'erreurs. Les performances de ces différentes associations de méthodes de normalisation avec les méthodes de combinaison sont présentées dans les tableaux 3.6 et 3.7.

Tableau 3.6: Performances des méthodes de normalisation qui ne modifient pas la forme des distributions associées aux méthodes de combinaison de scores simples.

Méthodes de normalisation	Taux de performance	Règles de Fusion					
		moyenne	produit	minimum	maximum	médiane	Somme pondérée
Min-Max	TEE	1.95	2.03	3.11	2.04	2.18	1.95
	TFR	0.5	0.5	0.5	1	0.5	0.5
	TFA	2.8	2.99	4.49	2.93	15.36	2.8
	TV	96.7	96.51	95.01	96.07	84.14	96.7
Znorm	TEE	1.95	2.03	3	2.03	2.21	1.95
	TFR	0.5	0.5	1	0.5	0.5	0.5
	TFA	2.79	3.01	4.31	2.92	15.12	2.77
	TV	96.71	96.49	93.69	96.58	84.38	96.73
Tangente Hyperbolique "Tanh"	TEE	1.95	1.96	3	2.03	2.21	1.95
	TFR	0.5	0.5	1	0.5	0.5	0.5
	TFA	2.79	2.82	4.31	2.92	15.12	2.77
	TV	96.71	96.68	95.59	96.58	84.38	96.73
La médiane et l'écart absolu médian (MAD)	TEE	2	2.04	3.16	1.99	2.31	1.95
	TFR	0.5	0.5	1	0.5	0.5	0.5
	TFA	2.92	3.04	4.41	2.78	16.26	2.8
	TV	96.58	96.46	94.59	96.72	83.24	96.7

Dans le tableau 3.6, nous constatons qu'il n'y a pas de grandes différences de performances selon les normalisations et les méthodes de combinaison. En particulier, la méthode médiane qui donne un mauvais résultat pour les données de test.

La méthode somme donne de bons résultats pour les quatre normalisations et fait toujours mieux que le meilleur système. La somme pondérée est la meilleure méthode de combinaison. Elle consiste en la définition de poids associés à chaque système, poids calculés de façon à minimiser le taux d'erreur sur la base d'évaluation. Nous constatons que la somme pondérée améliore les performances comparée à la deuxième meilleure méthode de combinaison qui est la somme simple (avec des poids identiques pour tous les systèmes, ici $\frac{1}{4}$ sauf pour la normalisation Min-Max pour laquelle les performances sont équivalentes).

Chapitre 3 : Vérifications uni-modale et multi-algorithmes de visage 2D

Tableau 3.7: Performances des méthodes de normalisation qui modifient la forme des distributions associées aux méthodes de combinaison de scores simples.

Méthodes de normalisation	Taux de performance	Règles de Fusion					
		moyenne	produit	minimum	maximum	médiane	Somme pondérée
Fonction Quadratique-Linéaire-Quadratique	TEE	2.18	2.68	3.21	2.95	2.13	2.13
	TFR	0.5	0.5	0.5	1.5	0.5	0.5
	TFA	3.05	3.88	4.49	3.21	25.27	2.9
	TV	96.45	95.62	95.01	95.29	74.23	96.6
Fonction Double Sigmoidé	TEE	2.16	2.33	2.81	1.96	2.15	2.17
	TFR	0.5	0.5	1	0.5	0.5	0.5
	TFA	3.15	3.5	4.07	2.74	14.03	3.14
	TV	96.35	96	94.93	96.76	85.47	96.36

Dans le tableau 3.7, nous constatons que les méthodes de normalisation associées aux méthodes de combinaison modifiant la forme des distributions n'améliorent pas la performance de vérification par rapport aux méthodes de normalisation qui ne modifient pas la forme des distributions. Mais nous aspirons à ce que ces méthodes donnent de bons résultats par la combinaison de scores par la logique floue.

Enfin nous pouvons affirmer que la somme et la somme pondérée sont les meilleures méthodes de combinaison simples. La somme pondérée étant légèrement meilleure mais nécessitant un apprentissage de poids.

- Comparaison des méthodes de normalisation par logique floue

Dans ce paragraphe, nous avons sélectionné trois méthodes de normalisation de scores:

- une méthode qui ne modifient pas la forme des distributions: c'est la Znorm (ayant donné le meilleur résultat par rapport à Min-Max, la médiane et l'écart absolu médian et c'est une méthode simple par rapport à celle de la tangente hyperbolique),
- deux méthodes modifiant la forme des distributions: la fonction Quadratique-Linéaire-Quadratique (QLQ) et fonction double sigmoïde.

Tableau 3.8: Performances des méthodes de normalisation par logique floue.

Méthodes de normalisation	Taux de performance	Règles de Fusion			
		Intégrale floue de Sugeno	B	Intégrale floue de Choquet	B
Znorm	TEE	2.08		2.16	
	TFR	0.5		0.5	
	TFA	3.22	0.3	3.19	0.1
	TV	96.28		96.31	
Fonction Quadratique-Linéaire-Quadratique	TEE	1.22		1.89	
	TFR	0.5		0.5	
	TFA	1.45	0.52	2.42	0.13
	TV	98.05		97.08	
Fonction Double Sigmoidé	TEE	1.14		2.4	
	TFR	0.5		0.5	
	TFA	1.14	0.50	3.82	0.12
	TV	98.36		95.68	

Le tableau 3.8 présente la comparaison des méthodes de normalisation associées aux méthodes de combinaisons par logique floue. Nous remarquons que la méthode de normalisation par la fonction

Chapitre 3 : Vérifications uni-modale et multi-algorithmes de visage 2D

Quadratique-Linéaire-Quadratique(QLQ) est intéressante pour les deux intégrales de *Sugeno* et *Choquet*. Nous obtenons une bonne performance avec la fonction double sigmoïde associée à l'intégrale de *Sugeno*.

Nous pouvons dire que la méthode de la fonction double sigmoïde nous donne le meilleur résultat pour l'intégrale floue de *Sugeno* avec un TEE=1.14 % et TV=98.36 %.

3.3.5.3 Méthodes de normalisation associées aux méthodes de classifications

Normalement les méthodes de fusion par classification de scores cherchent à prendre une décision à partir du vecteur de scores sans normalisation de scores. Dans notre cas, nous proposons d'utiliser les méthodes de normalisation de scores pour transformer individuellement chacun des scores afin de les rendre homogènes avant de les fusionner et ceci pour augmenter la performance d'authentification. Pour cela, nous utilisons trois méthodes de normalisation: Z-Score, fonction Quadratique-Linéaire-Quadratique (QLQ) et normalisation par une fonction double sigmoïde.

- Classification de scores par la méthode statistique de Fisher

Les différents taux d'erreur en utilisant la fusion par fonction linéaire discriminante (LDF) et fonction quadratique discriminante (QDF) sans et avec les méthodes de normalisation sont présentés dans le tableau 3.9.

Tableau 3.9: Performances des méthodes de normalisation par fusion statistique de *Fisher* dans l'ensemble d'évaluation et de test.

Méthodes de normalisation	Taux d'erreur	Règles de Fusion	
		LDF	QDF
sans normalisation	TEE	2.14	2.15
	TFR	0.5	0.5
	TFA	3.18	3.12
	TV	96.32	96.38
Znorm	TEE	1.98	2
	TFR	0.5	0.5
	TFA	2.84	2.89
	TV	96.66	96.61
fonction quadratique-linéaire-quadratique	TEE	1.97	2.
	TFR	0.5	0.5
	TFA	2.68	3.06
	TV	96.82	96.44
fonction double sigmoïde	TEE	2.17	2.18
	TFR	0.5	0.5
	TFA	3.15	3.19
	TV	96.35	96.31

Le tableau 3.9 montre clairement que les méthodes de normalisation améliorent la performance de l'authentification. On constate aussi qu'il n'y a pas de grandes différences de TEE pour les deux fonctions de fusion linéaire discriminante (LDF) et quadratique discriminante (QDF). Le meilleur résultat est obtenu (TEE = 1.97 % et TV= 96.82 %) par la méthode de normalisation de score avec la fonction Quadratique-Linéaire-Quadratique (QLQ) qui modifie la forme des distributions et la fonction de fusion LDF.

- Classification de scores avec les machines à vecteurs de support (SVM)

Nous utilisons une machine à vecteurs de support (SVM) avec un noyau RBF (Radial Basis Function ou fonction à base radiale) (voir équation 3.58).

La SVM a été implémentée en utilisant la librairie LibSVM disponible gratuitement au site (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>). Le noyau RBF utilisé est de la forme :

$$K_{RBF}(u, v) = e^{-\gamma \|u-v\|^2} \quad (3.63)$$

Où γ est un paramètre qui règle la largeur de la marge. La SVM nécessite également le réglage du paramètre C qui règle le compromis entre la largeur de la marge et le nombre d'erreurs admissibles (comme décrit dans l'équation 3.62).

Les différents taux d'erreur en utilisant la fusion par une machine à vecteurs de support (SVM) avec un noyau RBF sans et avec les méthodes de normalisation sont montrés dans le tableau 3.10.

Tableau 3.10: Performances des méthodes de normalisation associées à SVM dans d'ensemble d'évaluation et de test.

Règles de Fusion	Ensemble d'évaluation	Ensemble de test				
	TEE %	TFR %	TFA %	TV %	γ	C
Sans normalisation	2	0.5	2.93	96.53	1.5	0.6
Znorm	1.5	0.5	2.06	97.44	1.5	0.55
Fonction QLQ	1.96	0.5	2.14	97.36	2	0.7
Fonction double sigmoïde	1.66	0.5	2.39	97.11	1.6	3

Le tableau 3.10 montre clairement que les méthodes de normalisation améliorent la performance de l'authentification de visages. Notons également que le meilleur résultat avec un TEE=1.5 % et un TV=97.44% est obtenu par la méthode de normalisation Znorm qui ne modifie pas la forme des distributions.

- Classification de scores avec réseaux de neurones artificiels

Nous essayons finalement d'appliquer la fusion par classification de scores avec un réseau de neurones simple de type MLP (Multi Layer Perceptron). Les paramètres choisis pour notre expérience sont:

- Quatre entrées sont les scores pour chaque système.
- Une couche cachée avec dix neurones et la fonction d'activation sigmoïde.
- Deux neurones dans la couche de sortie et la fonction d'activation tangente hyperbolique.

Les différents taux d'erreur en utilisant un classifieur MLP sans et avec les méthodes de normalisation se trouvent dans le tableau 3.11.

Tableau 3.11: Performances des méthodes de normalisation associées à la classification MLP.

Règles de Fusion	Ensemble d'évaluation	Ensemble de test		
	TEE%	TFR %	TFA %	TV %
Sans normalisation	1.83	1	2.75	96.25
Znorm	1.69	1.5	1.81	96.69
Fonction QLQ	1.66	0.5	1.93	97.84
Fonction double sigmoïde	1.33	0.5	2.01	97.49

Le tableau 3.11 montre bien là aussi que les méthodes de normalisation améliorent la performance de l'authentification. Nous constatons qu'il n'y a pas de grandes différences de TEE selon les méthodes de normalisation. Celles qui modifient la forme des distributions donnent de bons résultats avec TEE=1.66 % et TV = 97.84 % par fonction Quadratique-Linéaire-Quadratique(QLQ) et TEE= 1.33% et TR = 97.49% par la fonction double sigmoïde.

Nous pouvons dire que la méthode proposée qui utilise la normalisation de scores avant la classification de scores améliore en général la performance d'authentification de visages. Dans la majorité des cas les méthodes de normalisation qui modifient la forme des distributions donnent de meilleurs résultats que les méthodes qui ne modifient pas la forme des distributions. Les deux méthodes de classification de scores SVM (TEE=1.5% et un TV=97.44%) et MLP (TEE= 1.33% et TV = 97.49%) donnent presque le même résultat.

La méthode retenue reste la fusion par combinaisons de scores par intégrale floue de *Sugeno* (TEE=1.14% et TV= 98.36%) qui donne le meilleur système multi-algorithmes d'authentification de visages avec un temps de calcul très réduit **t=0.94 seconde** (Langage de programmation utilisé est MATLAB R2009b et ordinateur: Intel Pentium Dual CPU 2.2 GHz, 1.49 GHz de RAM). Le meilleur système de fusion est donc celui qui est construit par la normalisation de scores par la fonction double sigmoïde et la fusion par combinaison par l'intégrale floue de *Sugeno*.

3.4 Conclusion

Ce chapitre est divisé en deux parties: vérification uni-modale et multi-algorithmes de visage 2D. Dans la première partie, nous avons tout d'abord présenté notre modèle de la vérification de visage. Puis, nous avons détaillé les quatre principales étapes de ce modèle qui sont: prétraitement, extraction des caractéristiques, réduction d'espace et comparaison. Enfin, l'étude expérimentale de l'authentification de visages sur la base de données XM2VTS est commencée par la comparaison des algorithmes de réduction d'espace, puis nous avons étudié l'extraction des caractéristiques par les ondelettes de *Gabor*.

Pour la comparaison de méthodes de réduction d'espace de données, nous avons utilisé quatre méthodes (Analyse en Composantes Principales (ACP), Analyse Discriminante Linéaire (LDA), le modèle discriminant Linéaire amélioré de *Fisher* (EFM) et LDA directe (DLDA)). Dans cette étude nous déduisons que la méthode EFM est la meilleure.

Pour l'extraction des caractéristiques, nous avons utilisé l'information spatio-fréquentielle du visage pour obtenir une bonne représentation. Cette information est extraite de l'image des pixels par application des 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 à cette limitation par un seuillage de la phase. Finalement, nous avons sélectionné les meilleurs systèmes d'authentification de visages pour construire un système multi-algorithmes dans la deuxième partie de notre chapitre.

Dans la deuxième partie de ce chapitre les limitations des systèmes biométriques lorsqu'ils n'utilisent qu'une seule modalité biométrique sont passées en revue. Ensuite, nous présentons une façon de réduire les limitations des systèmes biométriques monomodaux en combinant plusieurs systèmes menant ainsi à la biométrie multimodale. Les systèmes multimodaux peuvent être de différentes natures. Pour cela, nous étions amenés à étudier les différents types de combinaisons de modalités possibles, mais aussi les niveaux de fusion qui peuvent être utilisés dans un système multimodal. Nous présentons donc le type de multimodalités et le niveau de fusion. Parmi les divers niveaux de fusion existants, nous avons choisi de travailler au niveau score car il offre le meilleur compromis entre la richesse d'information et la facilité d'implémentation. Un état de l'art de la fusion de scores, la théorie des méthodes de fusion ainsi que six méthodes de normalisation de scores utilisées sont présentés.

Enfin, nous avons étudié dans les résultats expérimentaux un système multi-algorithmes d'authentification de visages. Celui-ci est construit par les quatre systèmes retenus dans la première partie de ce chapitre. Tout d'abord, nous avons présenté les distributions de scores et nous avons remarqué que les distributions sont différentes en termes de plages de variations et différentes dans leurs formes et leur recouvrement entre les deux classes (Client et Imposteur). Ce qui rend nécessaire l'étape de normalisation des scores. Nous avons comparé ensuite les méthodes de normalisation associées aux méthodes de combinaisons de scores simples. Nous pouvons dire que la somme et la somme pondérée sont les meilleures, la somme pondérée étant légèrement meilleure mais nécessitant un apprentissage des poids.

Pour la combinaison de scores par logique floue et en utilisant l'intégrale floue de *Sugeno* et *Choquet*, nous avons obtenu une bonne performance d'authentification de visage par l'intégrale floue de *Sugeno* et la méthode de normalisation de scores la fonction double sigmoïde avec un TEE=1.14 % et TV=98.36 %.

Dans la section de comparaison des méthodes de classification de scores nous proposons d'utiliser les méthodes de normalisation de scores avant la classification c'est ce qui fait l'originalité de notre travail. D'après toute cette étude et les résultats obtenus nous affirmons que les méthodes de normalisation de scores améliorent la performance en général pour toutes les méthodes de classification de scores utilisées. Les deux méthodes de classification de scores non linéaires SVM et MLP donnent des résultats équivalents. Cependant, le SVM présente une simplicité d'application.

Nous avons le long du chapitre travaillé uniquement sur la vérification de visage 2D. Malgré des résultats très intéressants, nous nous sommes rendus compte que nous atteignons des limites en ne considérant que la base de données contrôlée (XM2VST) qui ne contient pas de variation d'expression, changements en éclairage, rotation de la tête, bruit introduit par le capteur, occlusions Si l'on souhaite améliorer les performances et comme suite à notre travail nous pouvons proposer d'utiliser l'information 3D et la fusion des données 2D et 3D.

4.1 Introduction

La reconnaissance automatique des individus en fonction de leurs visages a de nombreuses applications utiles, en particulier dans des domaines tels que l'interaction homme-machine (robotique intelligente dans des environnements sociaux) et la sécurité (surveillance et contrôle d'accès). La nature non intrusive de reconnaissance basée sur la biométrie du visage, en comparaison des empreintes digitales et de l'iris, est plus appropriée pour un large éventail d'applications. Cependant, de nombreuses applications de sécurité nécessitent une précision de reconnaissance plus élevée que celle obtenue par les systèmes de reconnaissance actuels des visages. Ceci est vrai en particulier pour des variations d'expression du visage, d'éclairage et de pose [255]. Dans le chapitre 3, nous avons étudié la reconnaissance uni-modale puis multi-algorithmes de visage 2D. Malgré des résultats satisfaisants obtenus avec notre système d'authentification multi-algorithmes de visages qui est testé sur une base de données contrôlée (XM2VTS), les taux d'erreur associés à des systèmes biométriques restent relativement élevés lorsque les changements d'éclairage, d'expressions et de rotation de la tête sont pris en compte. Pour pallier à ces inconvénients, une solution est l'utilisation de l'image 3D. Ce chapitre présente donc la reconnaissance multimodale de visage par la fusion de l'information de profondeur (3D) et d'intensité (2D).

4.2 Le système de vérification multimodal proposé

Dans ce chapitre, nous examinons comment les caractéristiques locales de l'information 3D et 2D contribuent à la reconnaissance faciale lorsque l'éclairage, les changements d'expressions et la combinaison d'expression sous l'éclairage sont pris en compte. Tous les processus inclus dans la phase d'apprentissage et de test sont entièrement automatiques. Notre système, comme illustré sur la figure 4.1, comprend quatre étapes importantes:

- Prétraitement: par l'alignement (translation et rotation) de l'image d'entrée 3D par rapport à l'image de référence (image de visage d'expression neutre) en utilisant l'algorithme ICP (Iterative Closest Point). Selon les images alignées, nous obtenons des images de profondeur 3D et des images d'intensité 2D.
- Extraction des caractéristiques: une méthode d'extraction des caractéristiques robuste est très importante pour l'ensemble du système. Il est prévu que ces caractéristiques sont invariantes à la rotation, l'échelle, l'expression et l'éclairage. Dans notre système, nous utilisons les méthodes locales suivantes:
 - Multi-scale Local Binary Patterns (MLBP),
 - Caractéristiques Statistiques Locales (CSL),
 - Ondelettes de *Gabor*,
 - Transformation de caractéristiques visuelles invariantes à l'échelle (Scale-Invariant Feature Transform (SIFT)).
 - Descripteur BSIF (Binarized Statistical Image Features),
 - Descripteur LPQ (Local Phase Quantization).
- Classification: L'analyse en composantes principales (ACP) combinée avec le modèle discriminant linéaire amélioré de *Fisher* (EFM) sont utilisés pour la réduction et la classification des données. La comparaison est effectuée en utilisant le métrique cosinus (voir l'équation 3.24).
- Fusion: consiste en la fusion des résultats de la classification par la méthode SVM (Support Vector Machines) et en la normalisation de scores en utilisant la méthode Min-Max qui est choisie pour sa simplicité.

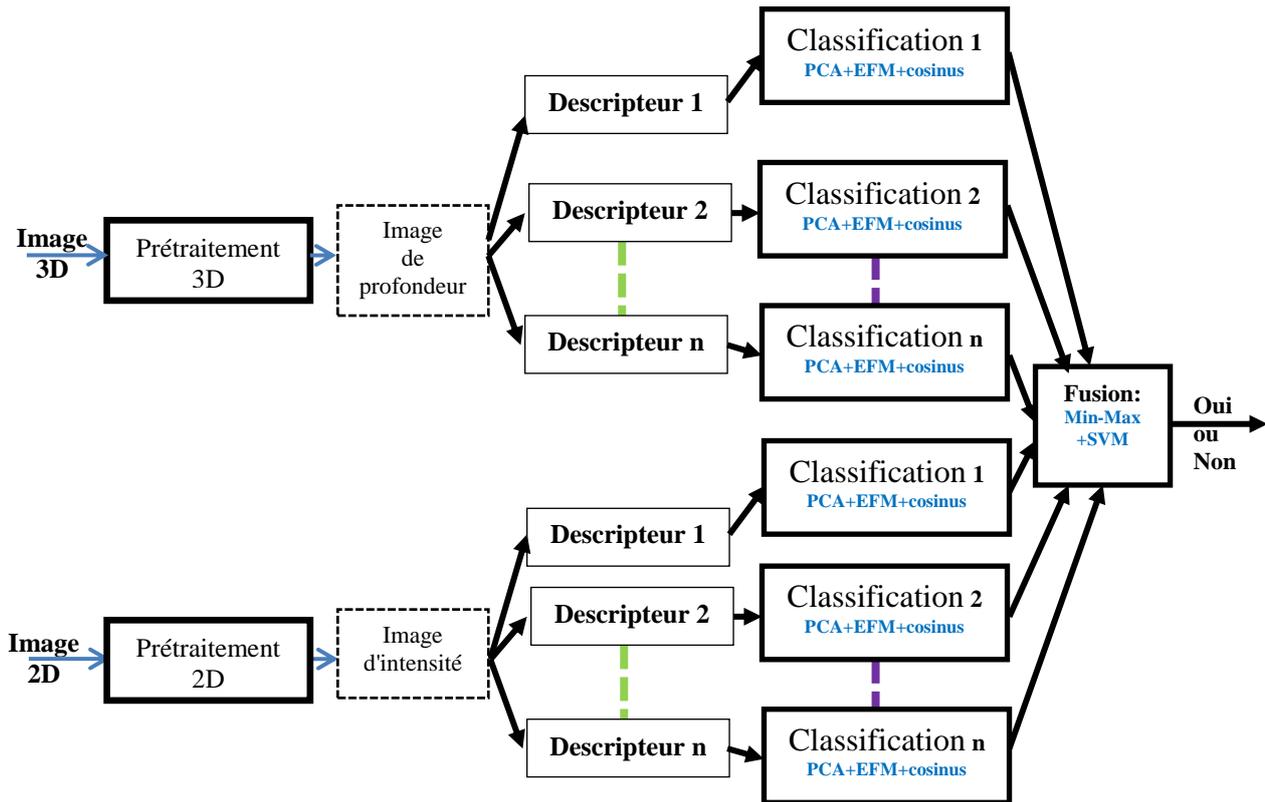


Figure 4.1: Vue de notre système proposé.

4.2.1 Prétraitement

On suppose qu'un visage est décrit par un nuage de points 3D capturés par un scanner laser 3D. Chaque nuage de points est constitué de milliers de points dans l'espace 3D. Ces points discrets décrivent approximativement la surface de la face. Nous utilisons la base de données de visage CASIA 3D. Chaque point est décrit avec des coordonnées spatiales 3D et les coordonnées de couleurs RGB correspondantes. Dans cette section, nous expliquerons comment les données 3D originales sont prétraitées.

Nous alignerons les données, puis nous obtiendrons les images de profondeur et d'intensité. Cette partie prépare l'extraction de caractéristiques dans la section suivante. Ce prétraitement comporte deux étapes principales, l'alignement des surfaces de visage 3D et l'acquisition de la profondeur et de l'intensité des images.

4.2.1.1 Alignement

L'alignement (ou le recalage) d'entités 3D consiste à trouver un ensemble de transformations en vue de les rapprocher. Ce sujet a suscité l'intérêt des chercheurs, depuis une vingtaine d'années, compte tenu de ses nombreuses applications. Le progrès le plus significatif a eu lieu au début des années 90 avec les travaux très similaires de *Besl et al.* [313] et *Chen et al.* [314] qui décrivent deux techniques itératives d'alignement. Nous utilisons l'algorithme du plus proche voisin itéré ICP (Iterative Closest Point) introduit par *Besl et al.* en 1992 dans [313]. ICP permet de trouver la transformation rigide $(R; t)$ qui minimise la distance entre deux ensembles de points $P = \{p_i\}$ et $X = \{y_i\}$. Le principe consiste à déterminer, pour chaque point p_i de l'ensemble P , le point le plus proche dans le second ensemble X au sens d'une distance euclidienne. A partir de ces appariements, une transformation rigide minimisant un critère des moindres carrés (4.1) est calculée et appliquée à l'ensemble P :

$$e(R, t) = \frac{1}{N} \sum_{i=0}^N \|(Rp_i + t) - y_i\|^2 \quad (4.1)$$

Ce processus est itéré jusqu'à la convergence caractérisée par une stabilité de l'erreur minimale par exemple. Dans ce processus, la transformation globale $(R; t)$ est mise à jour de façon incrémentale comme suit: pour chaque itération k de l'algorithme, $R = R_k:R$ et $t = t + t_k$. Le critère à minimiser à l'itération k devient alors (4.2) :

$$e(R_k, t_k) = \frac{1}{N} \sum_{i=0}^N \|(R_k(Rp_i + t) + t_k - y_i)\|^2 \quad (4.2)$$

L'algorithme ICP peut être résumé par les étapes suivantes :

- 1) On initialise l'algorithme pour $k = 0$ et on pose $S^0 = X = \{y_i\}$.
- 2) A l'itération k , on cherche pour chaque point p_i^k de P_i^k , le point de la première entité géométrique qui lui est le plus proche.
- 3) La meilleure transformation rigide qui recale les deux ensembles appariés est estimée à l'aide de l'une des méthodes d'optimisation de transformation rigide.
- 4) L'ensemble des points S^{k+1} est alors défini comme étant le transformé de l'ensemble S^k par la transformation rigide estimée à l'itération k .
- 5) Si le recalage effectué en 4) satisfait un critère de convergence, ce procédé est stoppé, sinon on réitère le même procédé en incrémentant k ($k = k + 1$) et en revenant à 2).

Dans notre cas la méthode ICP a deux attributions. Tout d'abord, ICP aligne toutes les images du visage par rapport l'image de référence (image du visage d'expression neutre). Deuxièmement, il examine si le bout du nez est détecté correct.

4.2.1.2 Image de profondeur et d'intensité

Les images de profondeur et d'intensité sont obtenues à partir des données 3D alignées. Les données sont converties en une image de profondeur 3D (voir figure 4.2 (a)) et une image en couleurs (voir figure 4.3 (a)). Dans la plupart des images, le nez est la partie du visage la plus proche au scanner 3D. Pour chaque pixel, la moyenne est calculée en utilisant une fenêtre de taille 9×9 autour de lui. L'utilisation d'une fenêtre de taille 3×3 qui calcule la somme de la profondeur de ses pixels correspondants, le nez est détecté en tant que coordonnées du pixel de la fenêtre centrale qui retourne la valeur minimale (voir figure 4.2 (b)). Après détection du nez, on extrait une sous-image centrée sur le centre de nez, avec une taille de 57×47 (voir figure 4.2 (c) et figure 4.3 (b)). Pour les images couleurs RVB, nous avons utilisé les images d'intensité (voir figure 4.3 (c)). Toutefois, en raison de la qualité originale des données 3D, les images de profondeur et d'intensité contiennent habituellement beaucoup de bruit, tels que des trous et des valeurs aberrantes. Nous pouvons obtenir des images améliorées par les procédés suivants:

Pour chaque pixel, la moyenne est calculée pour la fenêtre de 5×5 (voir figure 4.2 (d)) et figure 4.3 (d)). Si la valeur du pixel est inférieure à un seuil donné, ce pixel est remplacé par le pixel moyen. Le résultat est montré sur les figures 4.2 (e) et 4.3 (e). La variation de l'éclairage influe fortement sur la présentation des images d'intensité. Pour faire face à ce problème, l'égalisation d'histogramme est utilisée pour réduire l'influence des variations d'illumination (voir figure 4.3 (f)).

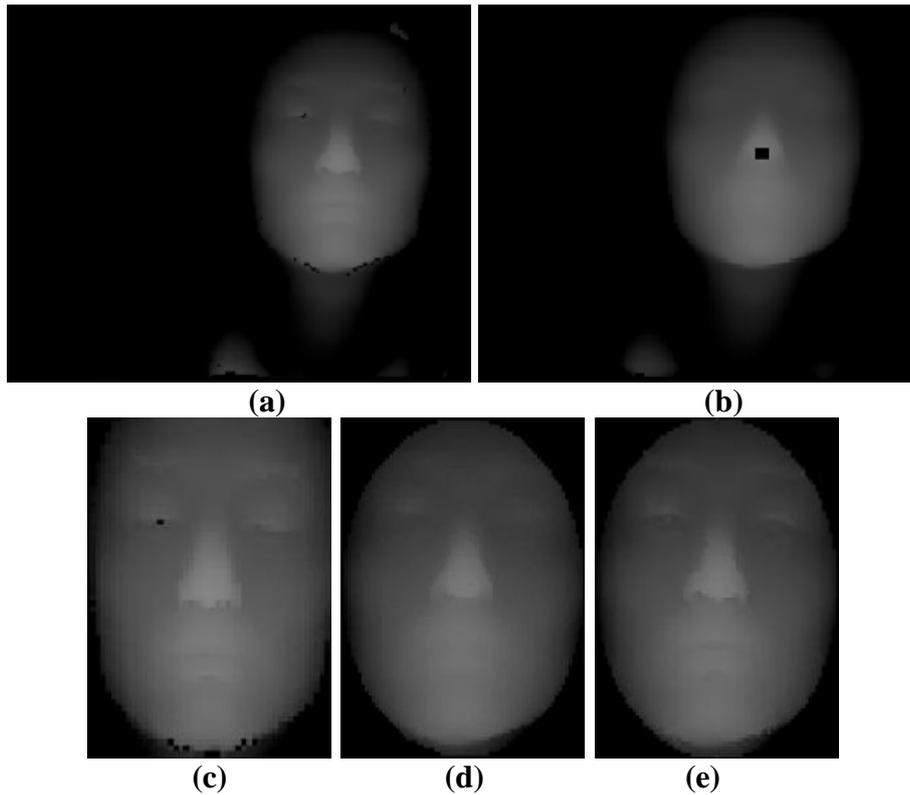


Figure 4.2: Prétraitement de l'image de profondeur.

(a) image de profondeur, (b) détection de la pointe du nez, (c) extraction du dessous-image, (d) image moyenne 5×5 , (e) image de profondeur après suppression du bruit et remplissage des trous.

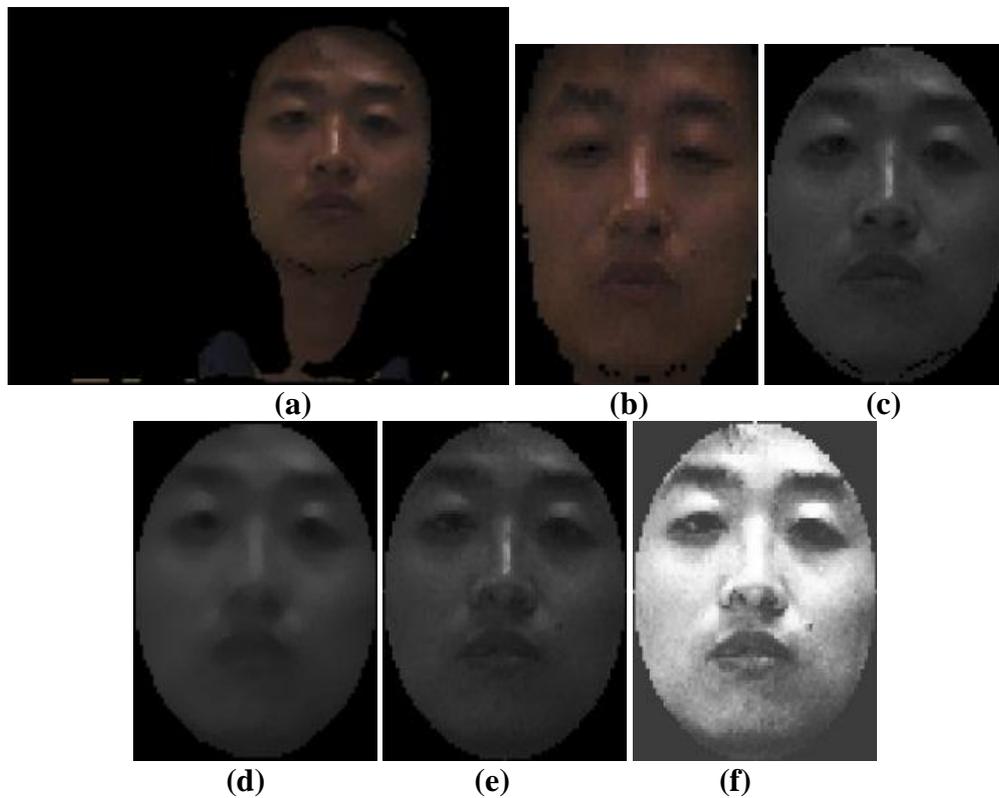


Figure 4.3: Prétraitement de l'image d'intensité

(a) image de la couleur, (b) extraction de sous-image, (c) image d'intensité, (d) image moyenne 5×5 , (e) image d'intensité après suppression du bruit et remplissage des trous, (f) image d'intensité après égalisation d'histogramme.

4.2.2 Extraction des caractéristiques du visage

L'information locale d'un visage est présente dans une région intérieure locale de l'image du visage, alors que l'information globale est extraite de toute l'image du visage. En général, les caractéristiques globales correspondent aux informations contenues dans les bandes des basses fréquences et les caractéristiques locales appartiennent à l'information dans les bandes des hautes fréquences qui dépendent de la position et de l'orientation des régions locales de l'image du visage. Dans ce chapitre, nous avons utilisé une approche globale basée sur la réduction d'espace de données, car elle reste parmi les techniques les plus utilisées et les plus populaires dans le domaine de la reconnaissance des visages. Dans ces procédés, les images du visage sont traitées de manière globale et sont généralement transformées en vecteurs, qui sont plus faciles à manipuler par les méthodes suivantes:

- Analyse en Composantes Principales (ACP),
- Analyse Discriminante Linéaire (ADL),
- Modèle discriminant Linéaire amélioré de *Fisher* (EFM),
- ADL directe (ADLD).

L'approche globale utilise dans son ensemble d'apprentissage une matrice. Ceci est vrai pour l'analyse discriminante linéaire à deux dimensions orientées en lignes (ADL2DOL) et en colonnes (ADL2DOC). Le but principal de ces méthodes est de trouver une base de projection qui optimise un certain critère. Cette base est ensuite projetée sur une base de dimension inférieure. Dans notre travail [315, 316], nous avons testé ces méthodes sur la base de données XM2VTS de visage 2D (voir chapitre 2) et la base de données CASIA du visage 3D. Nous avons montré que la méthode EFM est meilleure que d'autres méthodes d'extraction de caractéristiques.

Les méthodes d'extraction des caractéristiques locales étudiées dans le présent chapitre sont les suivantes:

- Local Binary Patterns (LBP): Récemment, la représentation LBP a gagné beaucoup d'attention dans le domaine de la reconnaissance faciale et est également appliquée avec succès en reconnaissance faciale 2D. Inspirés par l'opérateur LBP 2D, nous avons développé un algorithme de reconnaissance de visage 3D.
- Caractéristiques Statistiques Locales (CSL): calcule certains paramètres statistiques dans un voisinage de pixels.
- Les ondelettes de *Gabor* sont des descripteurs locaux les plus fréquemment utilisés dans la reconnaissance des visages. Les ondelettes de *Gabor* sont invariantes à l'échelle et l'orientation. Ces ondelettes sont capables de détecter des bords et des lignes dans les images du visage. Scale-Invariant Feature Transform (SIFT) se compose de caractéristiques invariantes à l'échelle et à la rotation.

Les paragraphes suivants présentent le calcul de ces caractéristiques.

4.2.2.1 Multi-scale local binary patterns (MLBP)

L'opérateur *LBP* a été proposé initialement par *Ojala et al.* [317] dans le but de caractériser la texture d'une image. Le calcul de la valeur *LBP* consiste pour chaque pixel à "seuiller" ses huit voisins avec un seuil dont la valeur est le niveau de gris du pixel courant. Tous les voisins prendront alors une valeur 1 si leur valeur est supérieure ou égale au pixel courant et 0 si leur valeur est inférieure (Fig.4.4). Le code *LBP* du pixel courant est alors produit en concaténant ces 8 valeurs pour former un code binaire. On obtient donc, comme pour une image à niveaux de gris, une image des valeurs *LBP* contenant des pixels dont l'intensité se situe entre 0 et 255.

La *LBP* a été étendue ultérieurement en utilisant des voisinages de taille différente. Dans ce cas, un cercle de rayon R autour du pixel central est considéré. Les valeurs des P points échantillonnés sur le bord de ce cercle sont prises et comparées avec la valeur du pixel central. Pour obtenir les valeurs des P points échantillonnés dans le voisinage pour tout rayon R , une interpolation est nécessaire. On

adopte la notation (P,R) pour définir le voisinage de P points de rayon R d'un pixel. La figure 4.5 (a) illustre trois voisinages pour des valeurs de R et P différentes.

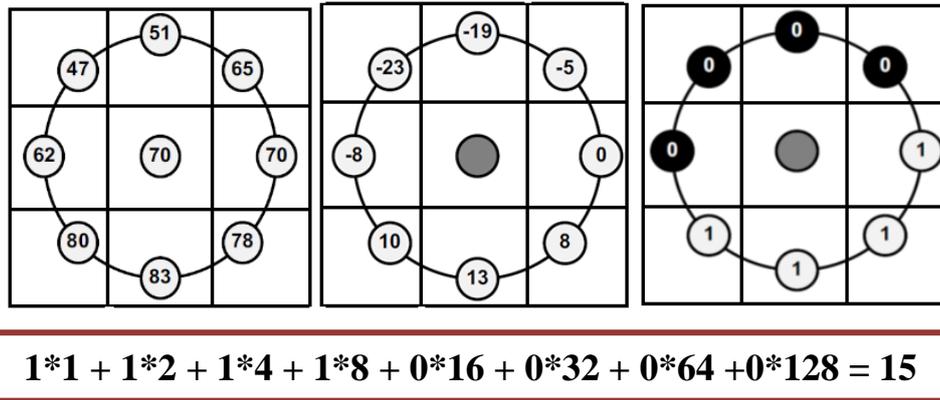
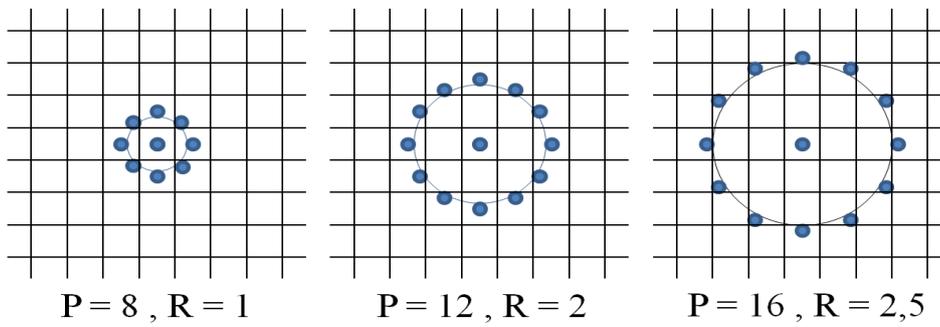
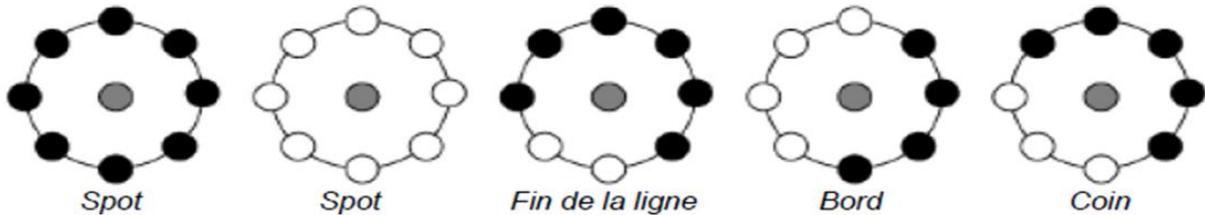


Figure 4.4: Opérateur LBP.



(a)



(b)

Figure 4.5: (a) Trois voisinages pour des R et P différents, (b) Textures particulières détectées par LBP

Soit g_c le niveau de gris du pixel central, g_p ($p = 1 \dots P$) les niveaux de gris de ses voisins, l'indice LBP du pixel courant est calculé comme :

$$LBP_{P,R}(x_c, y_c) = \sum_{p=1}^P s(g_p - g_c) 2^{p-1} \tag{4.3}$$

Où

$$s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} \tag{4.4}$$

Où (x_c, y_c) sont les coordonnées du pixel courant, $LBP_{P,R}$ est le code LBP pour le rayon R et le nombre de voisins P . L'opérateur LBP obtenu avec $P = 8$ et $R = 1$ ($LBP_{8,1}$) est très proche de l'opérateur LBP

d'origine. La principale différence est que les pixels doivent d'abord être interpolés pour obtenir les valeurs des points sur le cercle (voisinage circulaire au lieu de rectangulaire).

La propriété importante du code *LBP* est que ce code est invariant aux changements uniformes globaux d'illumination parce que la *LBP* d'un pixel ne dépend que des différences entre son niveau de gris et celui de ses voisins.

Compte tenu de l'image de profondeur et d'intensité du visage, nous générons un ensemble de multi-échelles *LBP* pour la représentation du visage. Quelques exemples sont illustrés dans la figure 4.6. Sur cette figure, le nombre de points d'échantillonnage varie de 8 points à 24 points et la valeur du rayon varie de 1 pixel à quatre pixels.

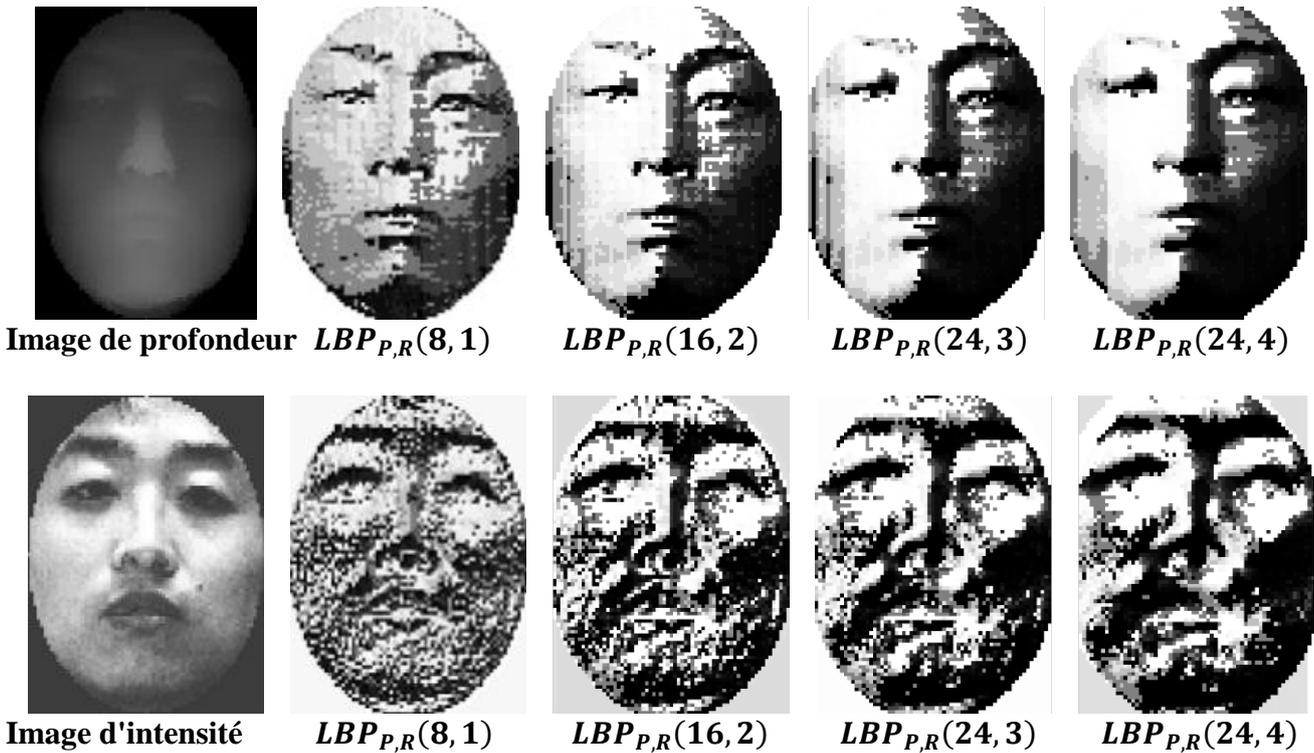


Figure 4.6: Multi-échelles *LBP* de l'image de profondeur et d'intensité du visage.

4.2.2.2 Caractéristiques statistiques locales proposées (CSL)

Le but principal de la méthode proposée est de calculer certains paramètres statistiques dans le voisinage du pixel en utilisant différentes tailles de R et le nombre de points voisins. Les paramètres calculés sont:

- **La moyenne (mean)**

Elle est définie comme:

$$mean_{P,R}(x_c, y_c) = \frac{1}{P} \sum_{p=0}^{P-1} i_p \quad (4.5)$$

Où i_c et i_p sont respectivement les valeurs du pixel central et P pixels voisins dans le cercle de rayon R .

- **Ecart type (standard deviation)**

L'écart-type montre combien il existe de variations par rapport à la moyenne. Il est défini comme:

$$std_{P,R}(x_c, y_c) = \sqrt{\frac{1}{P} \sum_{p=0}^{P-1} (i_p - mean_{P,R}(x_c, y_c))^2} \quad (4.6)$$

- **Variance**

Elle indique de quelle manière la variable aléatoire se disperse autour de sa moyenne. Une variance de zéro signale que toutes les valeurs sont identiques. Une petite variance est signe que les valeurs sont proches les unes des autres alors qu'une variance élevée est signe que celles-ci sont très écartées. Il est défini comme:

$$var_{P,R}(x_c, y_c) = \frac{1}{P} \sum_{p=0}^{P-1} (i_p - mean_{P,R}(x_c, y_c))^2 \quad (4.7)$$

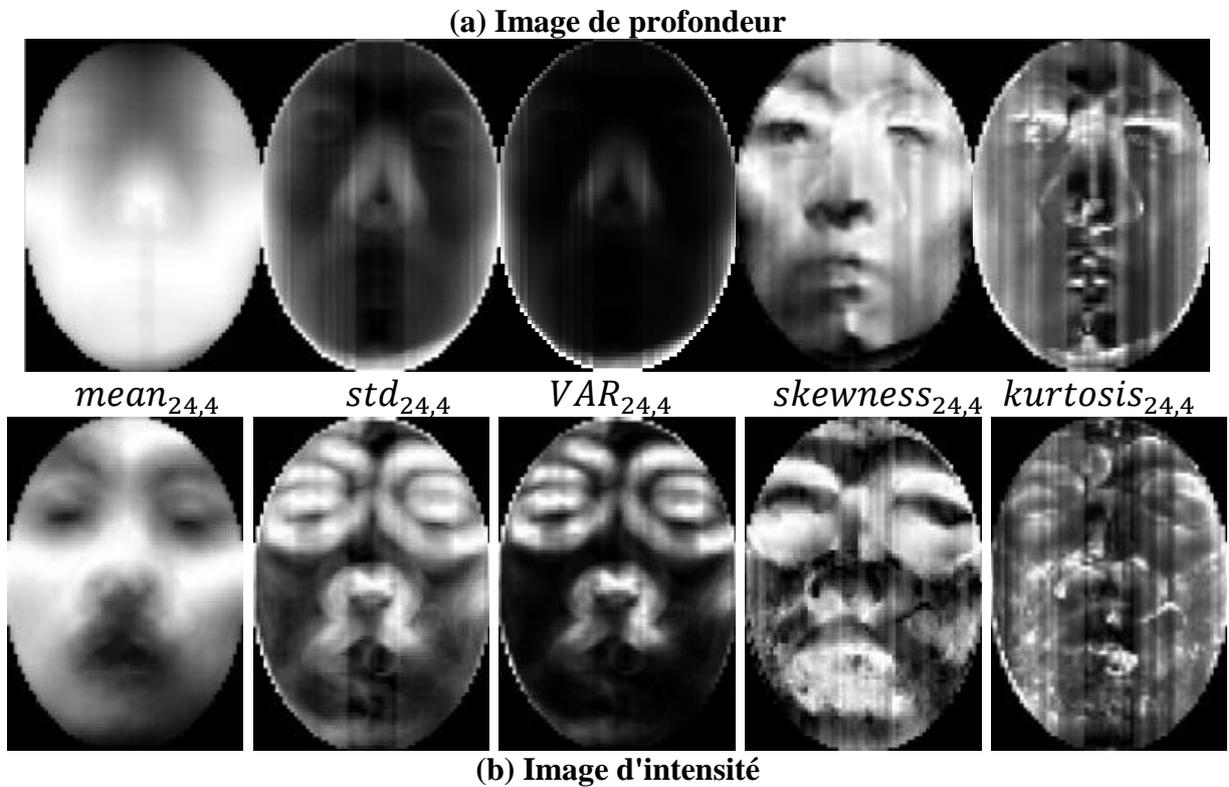


Figure 4.7: Caractéristiques statistiques locales proposées (CSL),
(a) image de profondeur, (b) image d'intensité.

- **Asymétrie (skewness)**

En termes généraux, l'asymétrie d'une distribution ou d'un ensemble de données est positive si la queue de droite (à valeurs hautes) est plus longue ou grosse et négative si la queue de gauche (à valeurs basses) est plus longue ou grosse. Elle est définie comme:

$$skewness_{P,R}(x_c, y_c) = \frac{\frac{1}{P} \sum_{p=0}^{P-1} (i_p - mean_{P,R}(x_c, y_c))^3}{(\sqrt{\frac{1}{P} \sum_{p=0}^{P-1} (i_p - mean_{P,R}(x_c, y_c))^2})^{3/2}} \quad (4.8)$$

- Aplatissement (kurtosis)

En théorie des probabilités et en statistique, le kurtosis, plus souvent traduit par coefficient d'aplatissement ou coefficient d'aplatissement de *Pearson*, correspond à une mesure de l'aplatissement ou inversement de la "pointicité" de la distribution d'une variable aléatoire réelle. Il mesure, hors effet de dispersion (donnée par l'écart type), la disposition des masses de probabilité autour de leur centre, c'est-à-dire, d'une certaine façon, leur regroupement proche ou éloigné du centre de probabilité. L'aplatissement est défini comme:

$$kurtosis_{P,R}(x_c, y_c) = \frac{\frac{1}{P} \sum_{p=0}^{P-1} (i_p - mean_{P,R}(x_c, y_c))^4}{(\frac{1}{P} \sum_{p=0}^{P-1} (i_p - mean_{P,R}(x_c, y_c))^2)^2} \quad (4.9)$$

La figure 4.7 montre les image de notre descripteur proposer (caractéristiques statistiques locales) pour les image de profondeur et d'intensité ($P=24, R=4$).

4.2.2.3 Ondelettes de *Gabor*

Dans ce chapitre, nous utilisons les filtres de *Gabor* sur les images de profondeur et d'intensité pour caractériser les visages. La représentation des visages à l'aide des ondelettes de *Gabor* a été utilisée avec succès pour la reconnaissance de visage en 2D et 3D [253]. Cette représentation d'une image décrit les caractéristiques faciales de la fréquence spatiale et les relations spatiales [318]. Pour plus de détails, on peut se référer au chapitre 3.

4.2.2.4 Scale-Invariant Feature Transform (SIFT)

Scale-Invariant Feature Transform (SIFT), que l'on peut traduire par « transformation de caractéristiques visuelles invariante à l'échelle », est une approche pour la détection et l'extraction de descripteurs de caractéristiques locales qui sont raisonnablement invariantes aux changements dans l'éclairage, le bruit d'image, la rotation, la mise à l'échelle et de petits changements de points de vue. Il a été développé en 1999 par le chercheur *David Lowe* [319]. Selon [319], la base de l'algorithme SIFT se compose de cinq étapes:

- détection d'extrema dans l'espace des échelles,
- localisation précise de points clés,
- affectation d'orientation,
- calcul de descripteurs de points-clés,
- correspondance.

Nous avons appliqué l'opérateur SIFT sur les images $LBP_{P,R}$ séparément. Parce que la $LBP_{P,R}$ met en évidence les caractéristiques locales lisses de l'image de profondeur et l'intensité du visage. On peut mieux détecter un nombre de points-clés par l'utilisation de SIFT sur l'image $LBP_{P,R}$ que l'utilisation de SIFT dans les images originales. Nous avons effectué un travail statistique sur la base de données CASIA3D. Le nombre moyen de points-clés extraits de chaque image $LBP_{P,R}$, est égal à 52 pour les images de profondeur et 162 pour les images d'intensité. En revanche, le nombre moyen de points-clés extraits de chaque image faciale d'origine de profondeur est limité à 14. Pour l'intensité, le nombre de points-clés est limité à 63. Nous avons utilisé le code *Lowes* [321] pour calculer le descripteur.

La figure 4.8 montre les points-clés extraits sur l'image du visage originale et de ses quatre images $LBP_{P,R}$ associés.

Pour calculer la similarité entre le visage d'apprentissage et de test, les points-clés du descripteur SIFT ont été appariés en utilisant la distance euclidienne (voir figure 4.9).

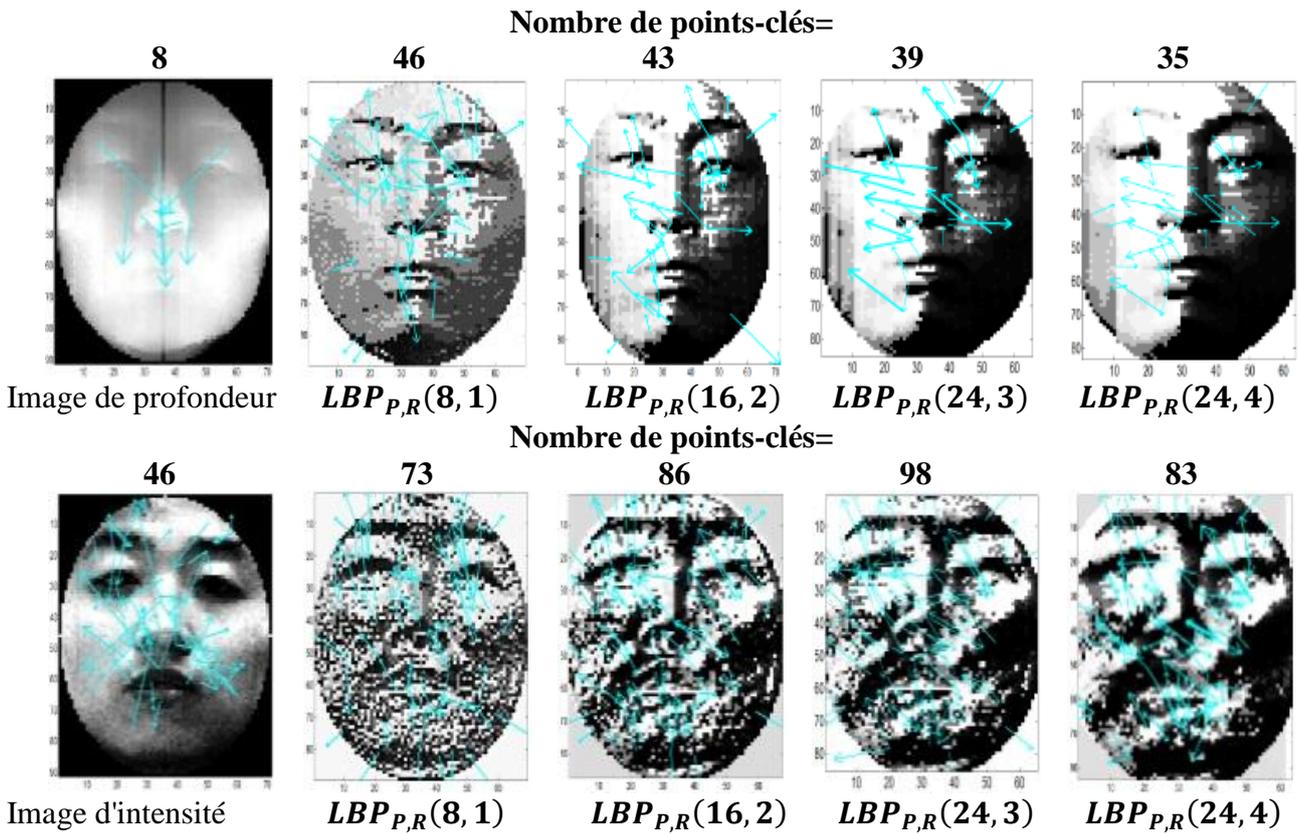


Figure 4.8: Points-clés détectés avec le descripteur SIFT de l'image de profondeur et d'intensité originale et les quatre images $LBP_{P,R}$ associées

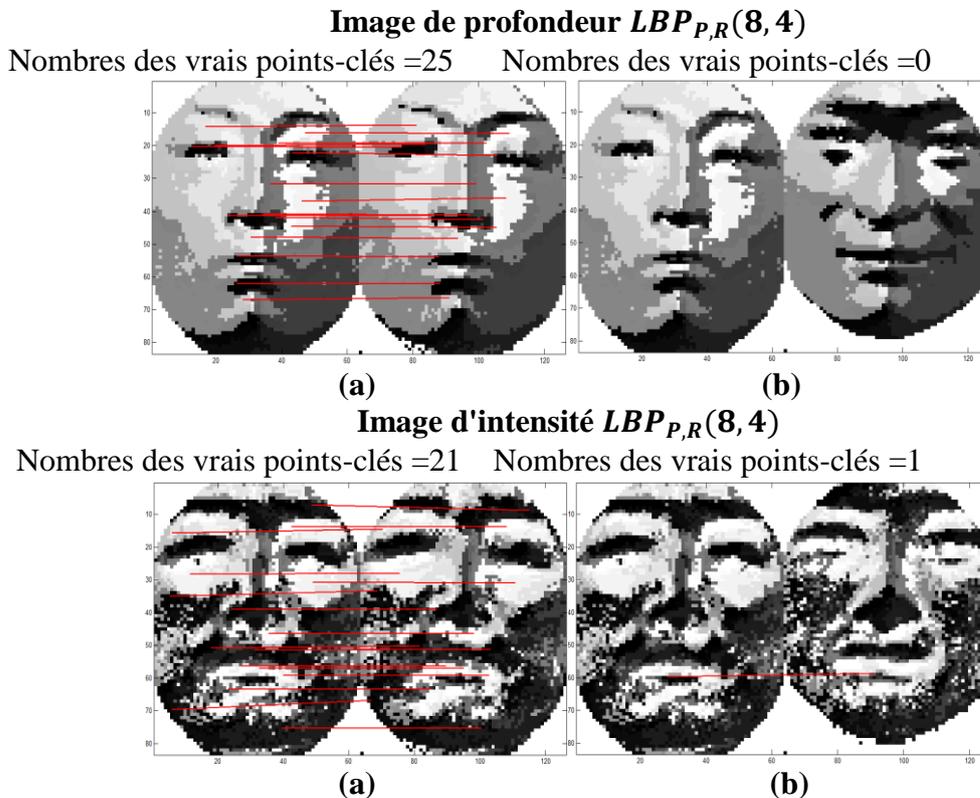


Figure 4.9: SIFT correspondante entre les visages d'apprentissage et de test appartenant à (a) la même identité et (b) des identités différentes.

4.2.3 Scores Fusion

Un système de fusion est constitué de deux modules, un module de fusion et un module de décision (voir figure 4.10). Le problème devient donc un problème de classification à 2 classes (OUI ou NON, Client ou Imposteur) à partir d'un vecteur de nombre réels dont la dimension est égale au nombre de sous-systèmes.

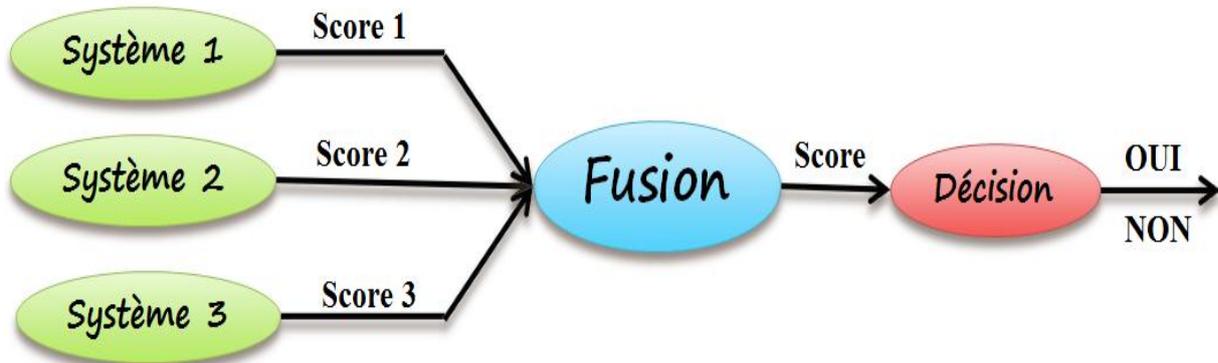


Figure 4.10: Schéma de la fusion des scores.

Il existe deux approches pour combiner les scores obtenus par différents systèmes:

- i) La première approche est de traiter le sujet comme un problème de **combinaison**,
- ii) La deuxième approche est de voir cela comme un problème de **classification**.

Dans nos travaux [316, 318], nous avons étudié deux approches pour la fusion de scores et plusieurs méthodes de normalisation des scores (voir chapitre 3), mais dans ce chapitre, nous avons utilisé comme normalisation de scores la méthode Min-Max (équation 3.27) pour sa simplification. Pour la fusion, nous avons utilisé une méthode de classification appelée Support Vector machines (SVM) (voire section 3.3.4.3), car c'est un procédé bien connu dans le domaine de la classification à deux classes.

4.2.4 Résultats

4.2.4.1 Base de données CASIA 3D

Entre Août et Septembre 2004, une base de données de visage 3D constituée de 4624 scans de 123 personnes est construite par l'utilisation d'un scanner 3D sans contact (Minolta Vivid 910), comme le montre la figure.4.11. Pendant la construction de la base de données, on a considéré non seulement les variations uniques de poses, les expressions et les illuminations, mais aussi les variations combinées des expressions sous éclairage et pose sous expressions, comme le montrent les figures 4.12, 4.13 et 4.16. Pour les sujets avec des lunettes, on a recueilli une acquisition supplémentaire avec des lunettes. Ainsi, chaque personne dispose de 37 ou 38 balayages. A partir de chaque balayage, une image couleur 2D et une surface triangulaire 3D du visage sont également générées. On a visé à construire une base de données complète du visage en 3D, qui est en outre amenée à être une plate-forme publique pour tester les algorithmes de reconnaissance faciale 3D ou autres [322].



Figure 4.11: Capture des scènes de la base de données CASIA visages 3D

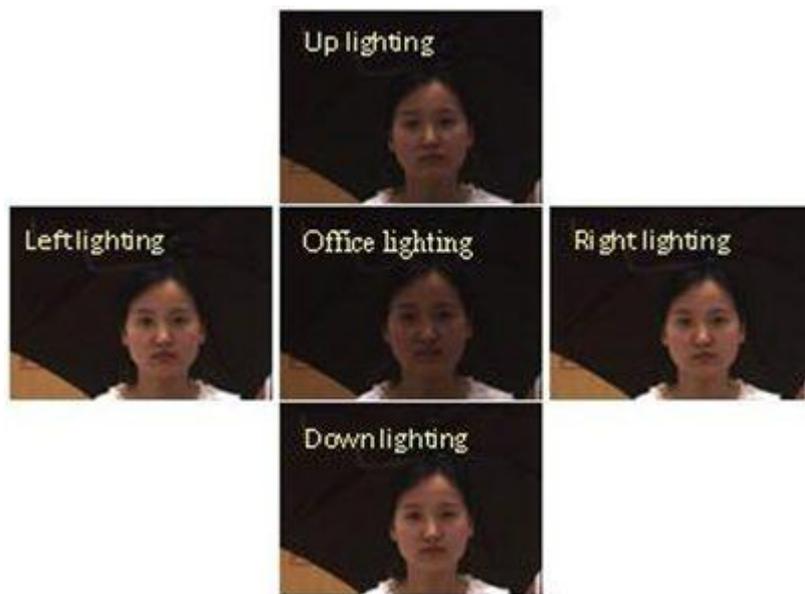


Figure 4.12: Variations d'éclairage de la base de données CASIA visage 3D.

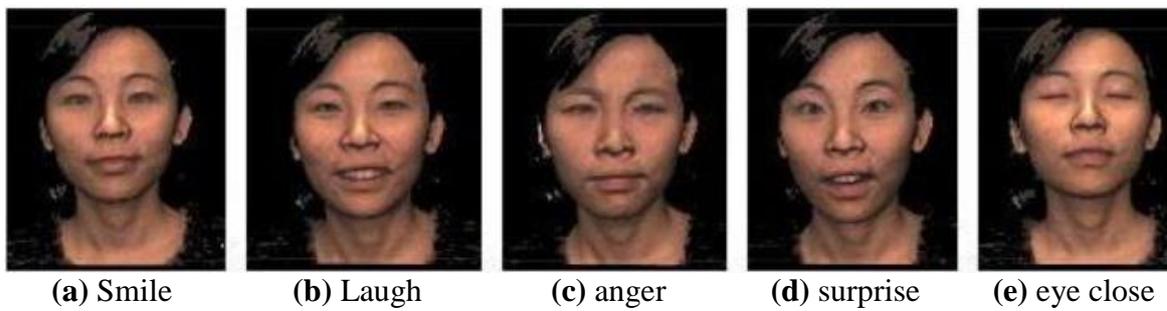


Figure 4.13: Variations d'expressions de la base de données CASIA visage 3D.

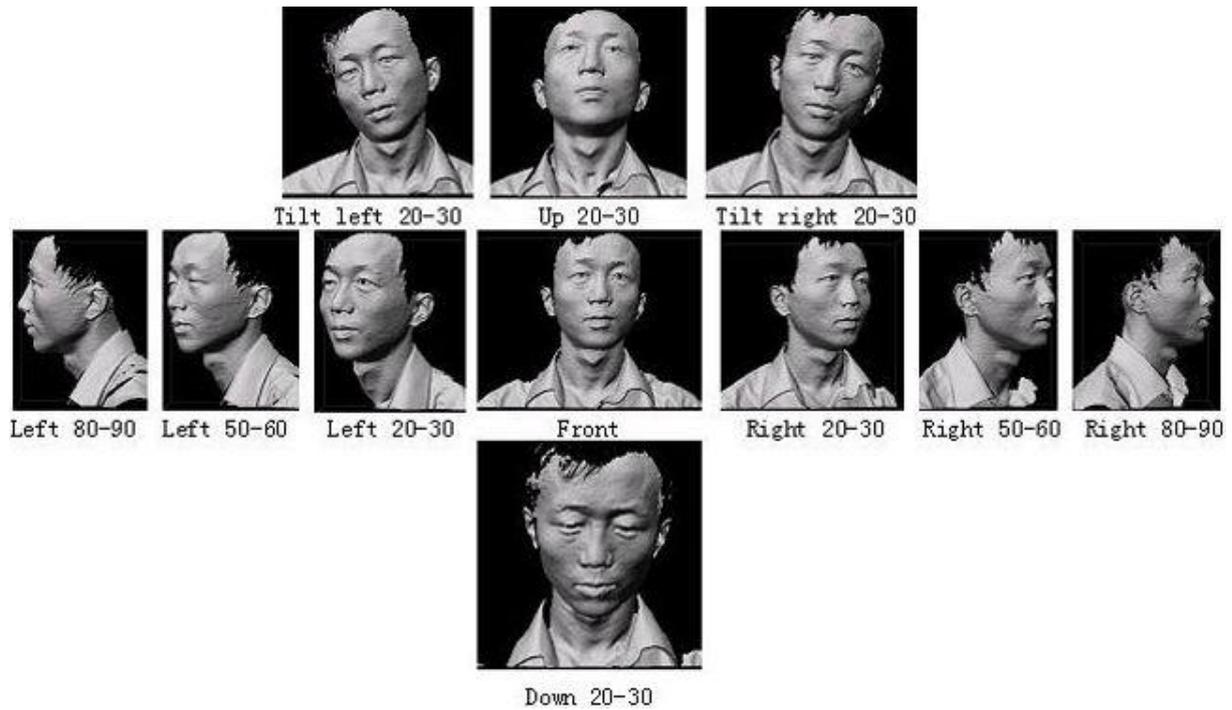


Figure 4.14: Variations des poses de la base de données CASIA visage 3D.

Dans cette section, nous avons étudié les variations d'éclairage (images: 1, 2, 3, 4, 5), expressions (images: 6, 7, 8, 9, 10) et les changements combinés dans l'expression sous éclairage (images: 11, 12, 13, 14, 15). Par conséquent, nous avons utilisé 15 images pour chaque personne. Nous avons utilisé un protocole d'évaluation pour séparer les personnes en deux classes, client et imposteur. Le groupe de clients contient 100 personnes, tandis que le groupe imposteurs est divisé en 13 imposteurs pour l'ensemble d'évaluation et de 10 pour l'ensemble de tests. La répartition des images dans différents ensembles est donnée dans le tableau 4.1.

Tableau 4.1: Répartition des photos dans différents ensembles

Ensemble	Client	Imposteur
Apprentissage	500 images (1, 4, 8, 9, 10)	0 images
Evaluation	500 images (2, 6, 7, 14, 15)	195 images (1:15)
Test	400 images (3, 5, 11, 12, 13)	150 images (1:15)

4.2.4.2 Approche globale (PCA + EFM)

Pour cette partie, nous utilisons une approche globale. Le vecteur caractéristique de l'image 2D et 3D est construit par concaténation des lignes de l'image de profondeur et d'intensité. Nous utilisons la méthode PCA + EFM pour la réduction et la séparation des données et de cosinus (expression 3.24) pour la mesure de similarité.

Le tableau 4.2 montre le taux d'erreur dans l'ensemble d'évaluation et de test d'une approche globale (PCA + EFM) d'extraction de caractéristiques.

- TEE: taux d'égale erreur,
- TV: taux de vérification ($TV = 100 - TFR - TFA$),
- TFR: taux de faux rejet,
- TFA: taux de fausse acceptation.

Chapitre 4: Vérification de visage 2D et 3D par CSL et les descripteurs locaux

NC est le nombre de caractéristiques extraites par l'EFM. Il est généralement calculé expérimentalement [316, 318] en faisant varier NC de 10, 20, ..., 200. Nous choisissons la valeur de NC qui donne le meilleur résultat.

Nous rappelons que P est le nombre de points voisins et R est le rayon du cercle contenant ces points voisins.

Le tableau 4.2 montre que PCA + EFM donne une mauvaise performance pour l'information de profondeur (3D). Nous avons obtenu: TV=89.36 % pour l'information de profondeur (3D), TV=92.26 % pour l'information d'intensité (2D) et TV=93.14 % pour la fusion multimodale (3D et 2D).

Tableau 4.2: Performance de PCA + EFM dans l'ensemble d'évaluation et de test.

Image de profondeur (2D)					Image d'intensité (3D)					Image de profondeur et d'intensité (2D et 3D)			
éval		test			éval		test			éval		test	
TEE	TFA	TFR	TV	NC	TEE	TFA	TFR	TV	NC	TEE	TFA	TFR	TV
3.36	2.94	4.8	92.26	30	7.24	4.24	6.4	89.36	10	2.61	2.86	4	93.14

4.2.4.3 Multi-échelle LBP (MLBP)

Dans cette partie, nous utilisons la méthode locale MLBP. Le tableau 4.3 montre le taux d'erreur dans l'ensemble d'évaluation et de test par cette méthode d'extraction de caractéristiques. Le nombre de points voisins (P) varie de 8 à 24 points et la valeur du rayon (R) varie de un à quatre pixels. La méthode LBP pour l'information 3D donne plus de performances que l'information 2D pour les quatre valeurs du rayon (R). La fusion des quatre valeurs du rayon (MLBP) améliore les performances des informations 2D et 3D et la fusion multi-capteurs (3D + 2D) avec un EER = 0,95 % dans l'ensemble d'évaluation et TV = 94,54 % dans l'ensemble de test.

Tableau 4.3: Performance de la méthode d'extraction de caractéristiques MLBP dans l'ensemble d'évaluation et de test.

(P, R)	Image de profondeur (2D)					Image d'intensité (3D)					Image de profondeur et d'intensité (2D et 3D)			
	éval		test			éval		test			éval		test	
	TEE	TFA	TFR	TV	NC	TEE	TFA	TFR	TV	NC	TEE	TFA	TFR	TV
1- (8,1)	5.90	6.93	7.4	85.66	60	4.74	3.11	5.2	91.68	50	3.82	4.26	3.6	92.14
2- (16,2)	4.22	5.41	5.2	89.38	100	6	5.79	3.8	90.40	50	3.56	4.08	3	92.92
3- (24,3)	4.37	5.16	4.8	90.04	70	6.17	5.79	4	90.2	50	3.79	4.08	2.9	93.12
4- (24,4)	4.17	6.58	4.4	89.02	100	6.17	5.52	4.2	90.27	50	3.61	4.35	3.2	92.44
1+2+3+4 (MLBP)	4.62	4.98	4.40	90.61	/	5.36	4.63	3.80	91.56	/	0.95	1.25	4.20	94.54

4.2.4.4 Caractéristiques statistiques locales (CSL)

Dans cette partie, nous utilisons la méthode locale proposée (CSL). Le tableau 4.4 montre le taux d'erreur dans l'ensemble d'évaluation et de test pour la méthode CSL. Le nombre de points voisins (P) varie de 8 à 24 points et la valeur du rayon (R) varie de un à quatre pixels. En premier lieu, la fusion des quatre rayons ($R = 1, 2, 3, 4$) pour les différents nombres de points voisins n'améliore pas les performances de cinq descripteurs statistiques disponibles. Nous remarquons aussi que, pour $R = 3, R = 4$ et le nombre de points $P = 24$ (taille maximale dans notre application), on obtient un meilleur résultat pour tous les descripteurs statistiques. Par conséquent, l'augmentation du nombre de points dans le voisinage améliore la performance dans le cas de descripteurs statistiques. Les quatre descripteurs (*moyenne, écart-type, variance, skewness*) donnent à peu près les mêmes résultats.

Kurtosis est le pire descripteur. Il confirme les résultats de la perception visuelle (qualité d'image) obtenus en images $kurtosis_{24,4}$ (voir figure 4.7).

La fusion des cinq paramètres de nos caractéristiques locales améliore l'authentification du visage. La performance sans *Kurtosis* est meilleure que la fusion des cinq paramètres statistiques avec TEE= 1,20 % dans l'ensemble d'évaluation et TV = 96,32 % dans l'ensemble de test pour la fusion multimodale (3D et 2D).

Tableau 4.4: Performance de la méthode d'extraction de caractéristiques CSL dans l'ensemble d'évaluation et de test.

(P, R)		Image de profondeur (2D)					Image d'intensité (3D)					Image de profondeur et d'intensité (2D et 3D)			
		éval	test				éval	test				éval	test		
			TEE	TFA	TFR	TV		NC	TEE	TFA	TFR		TV	NC	TEE
mean	1- (8,1)	3.36	2.73	4.4	92.86	50	6	3.89	5.4	90.70	10	2.17	1.74	3.6	94.66
	2- (16,2)	3.39	2.49	4.4	93.10	50	5.56	4.12	4.8	91.07	10	2.17	1.79	3	95.20
	3- (24,3)	3.23	1.99	5.2	92.80	30	5.20	4.78	4.8	90.41	10	1.95	0.96	3.6	95.43
	4- (24,4)	3.58	2.28	5.4	92.31	20	6	4.03	5.2	90.76	10	2.42	1.05	4.2	94.74
	1+2+3+4	3.18	2.41	3.8	93.78	/	5.18	3.96	5	91.04	/	1.3	0.68	4.2	95.11
std	1- (8,1)	3.76	3.3	4.8	91.89	60	11.59	14.39	10.4	75.20	30	3.57	3.36	4.2	92.16
	2- (16,2)	2.84	2.37	4.2	93.42	70	10.44	8.17	8.4	83.42	10	2.76	2.19	3.8	94.00
	3- (24,3)	3	2.13	3.2	94.66	80	9.24	7.79	7.6	84.60	20	1.96	2.4	2.8	94.80
	4- (24,4)	3.15	1.7	3.2	95.09	90	8.22	6.19	6.6	87.20	30	3.36	1.69	3	95.3
	1+2+3+4	3.04	2.41	2.6	94.98	/	5.63	4.86	7.4	87.74	/	1.63	1.26	3.6	95.13
var	1- (8,1)	5.61	4.76	7	88.23	70	11.8	8.62	11	80.38	20	4.62	4.22	4.6	91.17
	2- (16,2)	3.76	3.42	4.2	92.37	70	10.42	9.66	9.4	80.93	30	2.98	3.2	3.4	93.40
	3- (24,3)	2.97	2.48	3.4	94.11	70	11.19	11.34	10.2	78.46	40	2.81	2.68	3.2	94.11
	4- (24,4)	4.56	3.71	4	92.28	50	10.98	13.17	8.6	78.22	30	3.61	4.18	3.2	92.61
	1+2+3+4	3.36	2.84	3.8	93.35	/	10.21	10.24	8.2	81.55	/	1.40	1.6	4.4	93.99
skewness	1- (8,1)	5.61	4.26	5.2	90.53	80	11.80	5.29	5.4	89.30	80	3.79	3.58	4.4	92.02
	2- (16,2)	3.78	3.88	5.6	90.51	60	5.39	4.54	4.4	91.05	50	3.81	3.45	3.6	92.94
	3- (24,3)	3.38	3.3	4	92.70	90	4.82	4.46	4.8	90.74	60	3.40	3.2	3.2	93.60
	4- (24,4)	3	3.23	3.6	93.16	50	4.61	3.03	4.2	92.76	50	2.83	2.38	2.4	95.21
	1+2+3+4	2.96	3.06	4	92.94	/	5.16	4.28	4.4	91.31	/	1.61	1.72	3.8	94.48
kurtosis	1- (8,1)	18.78	20.33	18.6	61.06	20	14.84	14.08	11	74.91	20	12.98	13.51	14.2	72.28
	2- (16,2)	11.36	16.26	14.4	69.33	20	15.64	18.82	13.8	67.37	20	8.59	12.38	13.4	74.22
	3- (24,3)	10.18	15.46	12.4	72.13	40	20.44	23.30	20.6	56.10	20	9.16	14.74	13.4	71.86
	4- (24,4)	6.6	8.4	9	82.59	50	13.56	14.20	11.6	74.20	10	5.79	7.16	9.4	83.43
	1+2+3+4	4.76	7.11	15	77.88	/	11.59	13.28	10.6	76.12	/	5.16	7.24	8.4	84.35
mean(24,3)+std(24,4)+var(24,3)+skewness(24,4)+kurtosis(24,4)		4.98	2.14	2.60	95.25	/	3.60	2.06	3.80	94.13	/	0.99	0.80	4.00	95.20
mean(24,3)+std(24,3)+var(24,3)+skewness(24,4)		5.04	2.03	2.60	95.36	/	3.24	1.87	3.60	94.52	/	1.20	0.88	2.80	96.32

4.2.4.5 Ondelettes de Gabor

La famille des filtres de *Gabor* est caractérisée par un certain nombre de résolutions ou fréquences et orientations. Dans ce travail, nous avons concaténé pour chaque résolution les 8 directions dans un vecteur. Les filtres de *Gabor* ont une forme complexe qui peut être exploitée. Il est important d'utiliser les informations fournies par la partie réelle et la partie imaginaire des coefficients de *Gabor*. Dans nos travaux [318, 324], nous avons montré que la phase filtrée est plus performante que l'amplitude (voir la section 3.2.5.3). Nous utilisons donc les réponses de phase filtrées des filtres de *Gabor*.

Le tableau 4.5 montre le taux d'erreur dans l'ensemble d'évaluation et de test en utilisant les ondelettes de *Gabor* pour chaque résolution et la fusion de cinq résolutions. Les meilleurs résultats sont obtenus lorsque la résolution $\lambda = 4$, c'est-à-dire TEE = 1,57 % et TV = 96,13 %. La fusion de cinq résolutions n'améliore pas la performance de vérification de visages.

Tableau 4.5: Performance de la méthode d'extraction de caractéristiques par les ondelettes de *Gabor* dans l'ensemble d'évaluation et de test.

Résolution (λ)	Image de profondeur (2D)					Image d'intensité (3D)					Image de profondeur et d'intensité (2D et 3D)			
	éva TEE	test				éva EER	test				éva EER	test		
		TFA	TFR	TV	NC		TFA	TFR	TV	NC		TFA	TFR	TV
1-$\lambda=4$	2.56	1.18	3.40	95.42	40	5.41	4.08	4.20	91.71	20	1.57	0.66	3.2	96.13
2-$\lambda=4\sqrt{2}$	3.55	3.32	6	92.68	70	5.59	4.61	4.20	91.18	30	2.35	0.95	4.6	94.44
3-$\lambda=8$	4.55	2.42	8.2	89.38	40	7.77	5.57	5.20	89.22	20	3.00	1.51	5	93.48
4-$\lambda=8\sqrt{2}$	7.58	3.04	12	84.96	70	7.97	7.43	5.80	86.76	30	4.21	2	6.2	91.80
5-$\lambda=16$	7.2	3.82	10.2	85.97	30	4.96	5.98	5.2	88.81	10	3.83	1.88	5.2	92.92
1+2+3+4+5	2.20	1	4.6	94.40	/	3.55	3.06	4.2	92.74	/	0.84	0.38	5.2	94.41

4.2.4.6 MLBP + SIFT

Le tableau 4.6 indique le taux d'erreur dans l'ensemble d'évaluation et de test pour l'extraction de caractéristiques par MLBP + SIFT. Le nombre de points voisins (P) varie de 8 à 24 points et la valeur du rayon (R) varie de un à quatre pixels. Le tableau montre que la fusion de quatre LBP ($R=1, 2, 3, 4$ et $P=8, 16, 24$) plus SIFT donne le meilleur résultat avec un TEE =2,48 % et TV=94. 73 %.

Tableau 4.6: Performance de la méthode d'extraction de caractéristiques MLBP + SIFT dans l'ensemble d'évaluation et de test.

(P, R)	Image de profondeur (2D)				Image d'intensité (3D)				Image de profondeur et d'intensité (2D et 3D)			
	éval TEE	test			éval TEE	test			éval TEE	test		
		TFA	TFR	TV		TFA	TFR	TV		TFA	TFR	TV
1- (8, 1)	8.97	8.33	10	81.67	6.73	5.96	5.60	88.44	4.30	2.56	5	92.44
2- (16, 2)	6.09	5.33	7.2	87.47	6.30	3.11	7.00	89.89	5.31	5.62	3	91.37
3- (24, 3)	4.18	3.34	5	91.66	7.25	4.9	6.40	88.7	3.43	0.64	5	91.66
4- (24, 4)	3.7	5.11	5.4	89.49	6.34	4.89	4.60	90.51	2.85	2.3	4.2	93.50
1+2+3+4	2.91	2.46	4.6	92.94	4.67	2.41	4.6	92.98	2.48	2.26	3.00	94.73

4.2.4.7 Fusion des méthodes d'extraction de caractéristiques

Le tableau 4.7 compare les taux d'erreur dans l'ensemble d'évaluation et de test. Nous pouvons déduire que: le descripteur proposé CSL donne les meilleurs résultats avec TEE = 1,20 % dans l'ensemble d'évaluation et TV = 96,32 % dans l'ensemble de test. La fusion de notre descripteur CSL avec le descripteur MLBP +SIFT donne les meilleurs résultats. En effet, nous obtenons un TEE = 0,98 %, TV = 97,22 %. La fusion de tous les descripteurs considérés n'améliore pas la performance par rapport à la fusion de deux descripteurs CSL et MLBP +SIFT.

4.3 Extraction de caractéristiques par LPQ et BSIF

4.3.1 Extraction de caractéristiques par LPQ (Local Phase Quantization)

Le descripteur de texture LPQ a été introduit pour la première fois par *Ojansivu et al.* [152]. Il permet d'améliorer la classification de textures tout en étant robuste aux artéfacts générés par différentes formes de flou présents dans une image. Pour cela, le descripteur est construit de façon à ne retenir dans une image que l'information locale invariante à un certain type de flou. Les auteurs [152] ne considèrent en effet que les flous pouvant être représentés par une fonction d'étalement du point (*PSF*, "Point Spread Function") présentant une symétrie centrale. Cette hypothèse sur la *PSF* ne limite pas pour autant l'utilisation de cette méthode étant donné que la réponse à une source ponctuelle de la

majorité des capteurs et des systèmes d'imagerie peut être modélisée par ce type de fonctions mathématiques qui peuvent également présenter des symétries d'ordre supérieur (axiale ou radiale par exemple) [325]. Une fois les conditions sur le flou définies, une transformée de *Fourier* à fenêtre glissante est calculée pour plusieurs fréquences u choisies pour respecter les critères de la fonction d'étalement. Les coefficients ainsi obtenus sont quantifiés afin d'obtenir un mot de 8 bits [326].

Tableau 4.7: Performance de cinq méthodes d'extraction de caractéristiques dans l'ensemble d'évaluation et de test

extraction de caractéristiques	Image de profondeur (2D)				Image d'intensité (3D)				Image de profondeur et d'intensité (2D et 3D)			
	éval	test			éval	test			éval	test		
		TEE	TFA	TFR		TV	TEE	TFA		TFR	TV	TEE
1-PCA+EFM	3.36	2.94	4.8	92.26	7.24	4.24	6.4	89.36	2.61	2.86	4	93.14
2-MLBP	4.62	4.98	4.40	90.61	5.36	4.63	3.80	91.56	0.95	1.25	4.20	94.54
3-SLF	5.04	2.03	2.60	95.36	3.24	1.87	3.60	94.52	1.20	0.88	2.80	96.32
4-Gabort	2.56	1.18	3.40	95.42	5.41	4.08	4.20	91.71	1.57	0.66	3.2	96.13
5-LBP+SIFT	2.91	2.46	4.6	92.94	4.67	2.41	4.6	92.98	2.48	2.26	3.00	94.73
3+2	2.18	1.03	3.40	95.56	2.55	2.02	3.00	94.97	1.16	0.76	2.80	96.43
3+1	2.61	0.86	3.80	95.33	3.55	2.22	3.40	94.38	1.35	1.08	2.60	96.31
3+4	4.64	2	2.40	95.68	3.18	2	3.00	95.09	1.20	0.76	2.60	96.64
3+5	3.56	1.89	2.40	95.70	2.39	1.74	2.00	96.26	0.98	0.97	1.80	97.22
3+5+4	2.37	1.78	2.80	95.41	2.35	1.61	1.60	96.78	1.00	0.76	2.20	97.03
3+5+2	1.63	0.74	3.20	96.06	1.81	0.94	4.20	94.86	0.96	0.69	2.20	97.10
3+5+1	2.04	0.96	3.60	95.44	2.38	1.60	2.60	95.79	0.84	0.75	2.80	96.44
1+2+3+4+5	1.58	0.68	3.40	95.92	1.64	0.87	4.00	95.12	0.77	0.56	3.00	96.43

4.3.1.1 PLQ descripteur de base

L'information de LPQ peut être extraite en utilisant la transformée discrète de *Fourier* à fenêtre à deux dimensions (2DWFT).

$$F_u(x) = \sum_{m \in N_x} h(m-x)f(m)e^{-j2\pi u^T m} = E_u^T f_x \quad (4.10)$$

Où E_u , de taille $= 1 \times M^2$, est un vecteur de base de 2DWFT avec la fréquence u , et f_x , taille $= M^2 \times N$, est un vecteur contenant les valeurs des pixels d'image dans N_x à chaque position x . La fonction fenêtre, $h(x)$ est une fonction rectangulaire. La transformation est calculée à quatre valeurs de la fréquence, $u = [u_0, u_1, u_2, u_3]$ où $u_0 = [a, 0]^T$, $u_1 = [0, a]^T$, $u_2 = [a, a]^T$ et $u_3 = [a, -a]^T$. La valeur a est la plus haute fréquence scalaire pour laquelle $H_{u_i} > 0$. Ainsi, seuls quatre fonctions complexes comme un banc de filtres sont nécessaires pour produire huit images résultantes, composées de 4 images de la partie réelle et 4 images de la partie imaginaire de la transformée. Chaque pixel de l'image complexe résultant peut être codé en une valeur binaire représentée dans l'équation (4.11) en appliquant (the quadrant bit coding) [330].

$$B_{u_i}^{Re}(x) = \begin{cases} 1 & \text{si } F_{u_i}^{Re}((x)) > 0 \\ 0 & \text{si } F_{u_i}^{Re}((x)) \leq 0 \end{cases} \quad B_{u_i}^{Im}(x) = \begin{cases} 1 & \text{si } F_{u_i}^{Im}((x)) > 0 \\ 0 & \text{si } F_{u_i}^{Im}((x)) \leq 0 \end{cases} \quad (4.11)$$

Ce procédé de codage attribue deux bits pour chaque pixel pour représenter le quadrant dans lequel se trouve l'angle de phase [330]. En fait, il fournit également la quantification de la fonction de phase de *Fourier*. En général, LPQ est une chaîne binaire, présentée dans l'expression (4.12), obtenue pour chaque pixel par la concaténation des codes quadrant bits réelles et imaginaires des huit coefficients de *Fourier* de u_i .

$$LPQ(x) = [B_{u_0}^{Re}(x), B_{u_0}^{Im}(x), \dots, B_{u_3}^{Re}(x), B_{u_3}^{Im}(x)] \quad (4.12)$$

La chaîne binaire est convertie en nombre décimal par l'expression (4.13) pour produire une étiquette de LPQ. La figure 4.15 résume l'ensemble de ces étapes.

$$LPQ(x) = B_{u_0}^{Re}(x) + B_{u_0}^{Im}(x) \times 2^1 + \dots + B_{u_3}^{Re}(x) \times 2^{k-1} + B_{u_3}^{Im}(x) \times 2^k \quad (1.23)$$

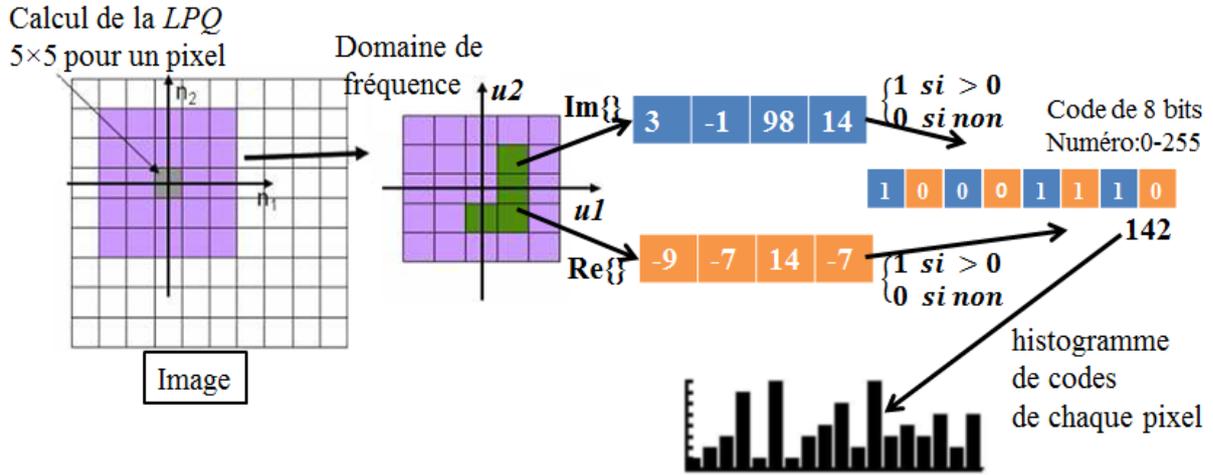


Figure 4.15: Organigramme de l'ensemble des étapes nécessaire à la construction du descripteur LPQ.

4.3.1.2 Processus de décorrélation

En général, les valeurs des pixels adjacents dans une image capturée par une caméra sont fortement corrélées, ce qui conduira à une dépendance entre $F_u(x)$ quantifiée en $LPQ(x)$. Ainsi, *Ojansivu et al.* [152] ont proposé de décorréler $F(x)$ par une transformée de blanchiment représentée par l'expression (4.14) en tant qu'information dans la quantification qui est au maximum conservée dans l'échantillon statistiquement indépendant.

$$Q(x) = W^T F(x) \quad (4.14)$$

Où W est la matrice des vecteurs propres dérivés par la fonction de décomposition propre de la matrice de covariance D .

Le processus de décorrélation améliore la performance du système. La matrice D estimée par l'expression (4.15) est la matrice de covariance des coefficients de *Fourier* $F(x)$.

$$D = ECE^T, \quad E = [E_{u_0}^{Re}, E_{u_0}^{Im}, E_{u_1}^{Re}, E_{u_1}^{Im}, E_{u_2}^{Re}, E_{u_2}^{Im}, E_{u_3}^{Re}, E_{u_3}^{Im}]^T \quad (4.15)$$

Où $E_{u_i}^{Re}$ et $E_{u_i}^{Im}$ sont les parties réelles et imaginaires du vecteur de base de 2DWFT à la fréquence de u_i , et C est la matrice de covariance.

$$C = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \dots & \sigma_{1,M^2} \\ \sigma_{2,1} & \sigma_{2,2} & \dots & \sigma_{2,M^2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M^2,1} & \sigma_{M^2,1} & \dots & \sigma_{M^2,M^2} \end{pmatrix} \quad (4.16)$$

Où σ_{ij} est la covariance entre les valeurs de pixels à des positions x_i et x_j dans N_x . Comme les valeurs des pixels adjacents sont fortement corrélés à l'image, il peut être supposé que les covariances peuvent

être modélisés sous la forme d'un processus de *Markov* de premier ordre, dans lequel le coefficient de corrélation entre deux valeurs de pixels peut être exprimé par:

$$\sigma_{i,j} = \rho^{\|x_i+x_j\|} \quad (4.17)$$

ρ étant compris entre 0 et 1, cela signifie que plus les deux pixels sont éloignés l'un de l'autre, plus la covariance diminue. Dans cette méthode, ρ est fixé à 0.95.

Dans la figure 4.16, une illustration de la représentation de l'image de profondeur et d'intensité pour le descripteur LPQ en utilisant cinq rayons ($R = 1, 2, 4, 6, 8$).

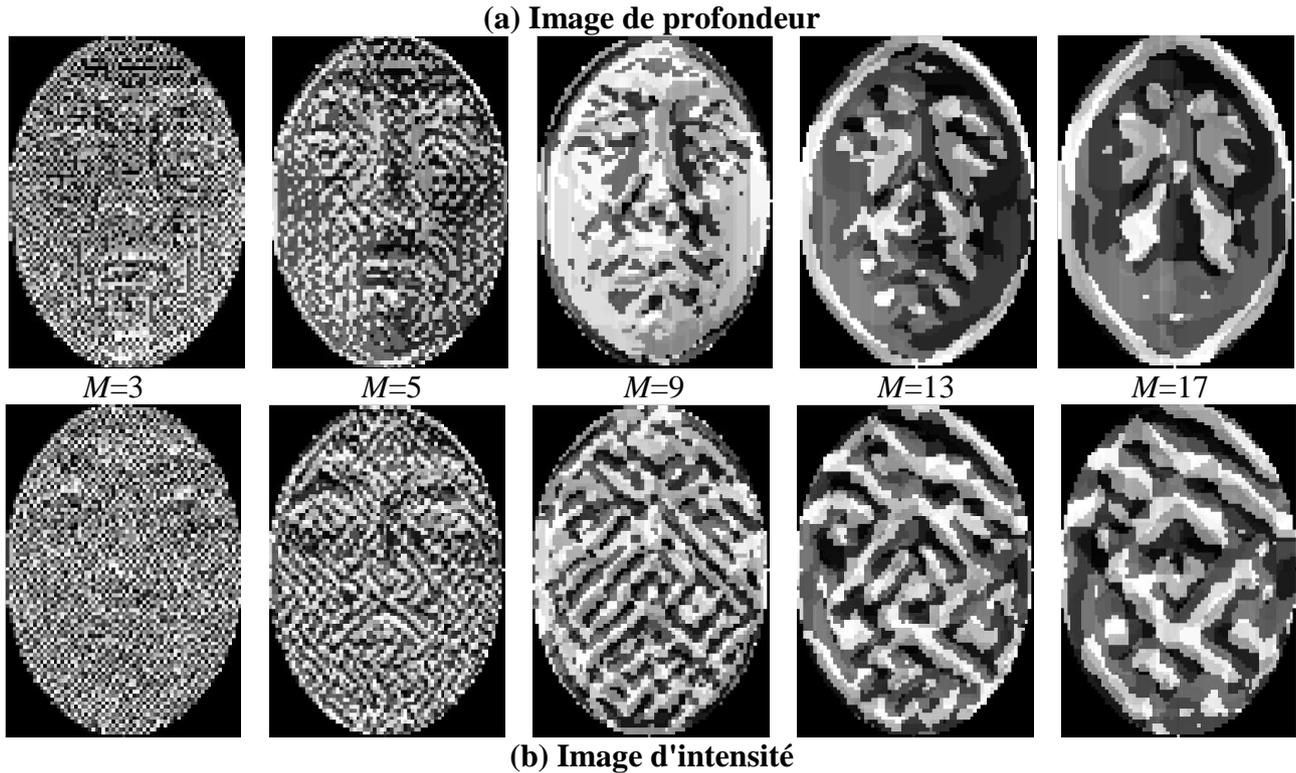


Figure 4.16: LPQ, (a) image de profondeur, (b) image d'intensité.

4.3.2 Extraction de caractéristiques par BSIF (Binarized Statistical Image Features)

Contrairement à la LBP et la PLQ qui peuvent être utilisées pour calculer les statistiques d'étiquettes dans les voisinages des pixels locaux, le nouveau descripteur local appelé BSIF (Binarized Statistical Image Features), qui a été récemment proposé par *Kannla and Rahtu* pour la reconnaissance faciale 2D et la classification de texture, utilise un ensemble prédéfini manuellement des filtres linéaires et binarisation des réponses du filtre [154].

4.3.2.1 Détails

Compte tenu d'une image patch X de taille $l \times l$ pixels et un filtre linéaire W_i de la même taille, la réponse du filtre s_i est obtenue par:

$$s_i = \sum_{u,v} W_i(u,v)X(u,v) = w_i^T x \quad (4.18)$$

Où la notation de vecteur est introduite dans la dernière étape, à savoir, les vecteurs w et x contiennent les pixels de W_i et X . La fonction binarisée b_i est obtenue par:

$$b_i = \begin{cases} 1 & s_i > 0 \\ 0 & \text{autrement} \end{cases} \quad (4.19)$$

Compte tenu de n filtres linéaires W_i , nous pouvons les empiler sur une matrice W de taille $n \times l^2$ et calculer toutes les réponses à la fois, c'est-à-dire $s = Wx$. On obtient la chaîne de bits b par la binarisation de chaque élément s_i de s comme ci-dessus. Ainsi, compte tenu des caractéristiques des détecteurs linéaires W_i , le calcul de la chaîne de bits b est simple. En outre, il est clair que les chaînes de bits pour tous les blocs de l'image de taille $l \times l$, entourant chaque pixel d'une image, peuvent être calculées facilement par n convolutions.

Afin d'obtenir un ensemble utile de filtres W_i *Kannla* et *Rahtu* [154] ont appliqué les concepts de [25] et estimé les filtres en maximisant l'indépendance statistique des s_i . Pour estimer les composantes indépendantes, il faut briser la matrice de filtre dans deux parties par:

$$s = Wx = UVx = Uz \quad (4.20)$$

Où $z = Vx$, et U est une matrice carrée de taille $n \times n$ qui sera estimée par ICA (Independent Component Analysis) et la matrice V effectue le prétraitement canonique, c'est le blanchiment et la réduction de la dimension simultanés des échantillons simple x [332].

En bref, le prétraitement canonique utilise l'analyse en composantes principales présentée comme ci-après.

Étant donné un ensemble d'images patchs sélectionnés au hasard à partir d'images naturelles de l'ensemble d'apprentissage, les patchs sont d'abord faits de moyenne nulle (l'intensité moyenne de chaque patch est soustraite) et leur dimension est réduite en ne conservant que les n premières composantes principales qui sont divisées par leur écart-type pour obtenir des échantillons de données blanchis z . Dans le détail, compte tenu de la décomposition en vecteurs propres et valeurs propres $C = EDE^T$ de la matrice de covariance C des échantillons x , la matrice V est définie par:

$$V = \left(D^{-\frac{1}{2}} E^T \right)_{1:n} \quad (4.21)$$

Où la diagonale principale de D contenant les valeurs propres de C dans l'ordre décroissant, et $(\cdot)_{1:n}$ désigne les n premières lignes de la matrice entre parenthèses.

Ensuite, compte tenu de la valeur moyenne nulle des échantillons de données blanchies z , on peut utiliser des algorithmes d'analyse en composantes indépendantes standards pour estimer une matrice orthogonale U avec laquelle on produit les composantes indépendantes s des données d'apprentissage [333]. En d'autres termes, étant donné que $z = U^1 s$, les composantes indépendantes permettent de représenter les échantillons z comme une superposition linéaire de vecteurs de base définis par les colonnes de U^1 de données. Enfin, compte tenu de U et V , on obtient la matrice de filtre $W=UV$, qui peut être directement utilisée pour le calcul BSIF.

4.3.2.2 Mise en œuvre

Dans toutes les expériences de *Kannla* et *Rahtu* [154], les filtres sont tirés d'un ensemble de 13 images naturelles fournies par les auteurs de [332] (voir figure 4.17). Avant l'échantillonnage aléatoire, des images patchs ont été normalisées pour avoir une moyenne zéro et une variance l'unité. Comme décrit ci-dessus, il existe deux paramètres dans le descripteur BSIF: la taille du filtre l et la longueur n de la chaîne de bits. Les filtres W sont construits par utilisation de plusieurs choix différents de valeurs de paramètres. Chaque ensemble de filtres a été appris par utilisation de 50 000 images patchs. L'apprentissage a été mené par le processus en trois étapes détaillées dans le paragraphe précédent: (a)

soustraction de l'intensité moyenne de chaque patch, (b) réduction de la dimension et le blanchiment par analyse en composante principale et (c) l'estimation des composantes indépendantes. Les filtres obtenus avec $l=7$ et $n=8$ sont illustrés par la figure 4.18.



Figure 4.17: 13 images naturelles utilisées pour la formation des Filtres W_i

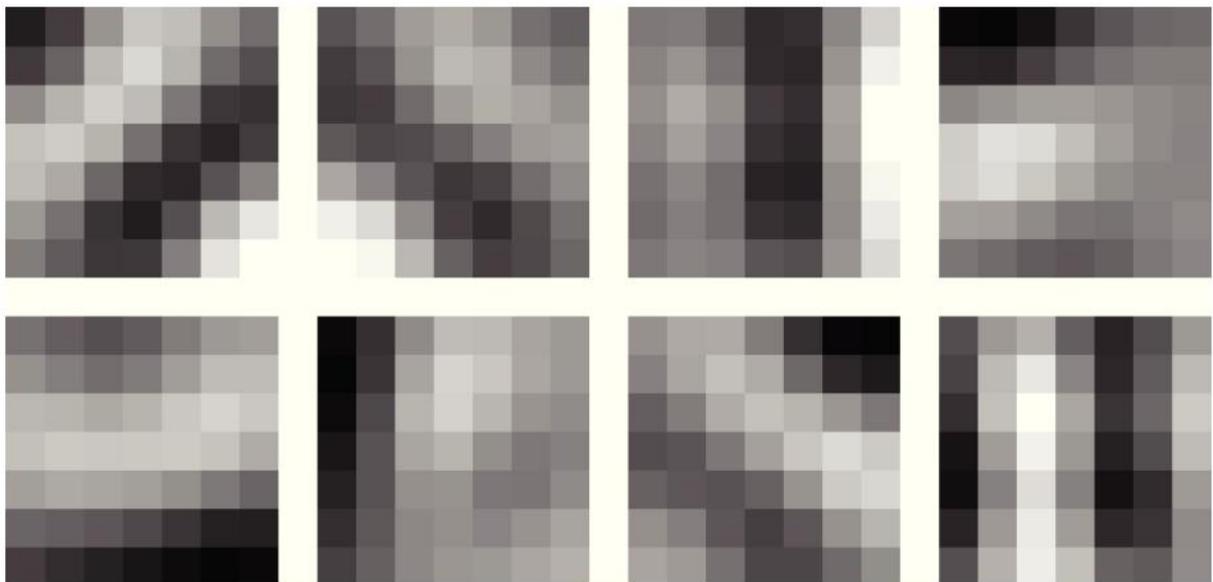


Figure 4.18: Filtres tirés de taille $l=7$ et nombre de bits $n=8$.

Comme décrit ci-dessus, il existe deux paramètres pour le descripteur BSIF: la taille du filtre l la longueur de la chaîne de bits n . Les images de profondeur de code BSIF et d'intensité correspondantes

sont présentées dans la figure 4.19. La taille du filtre l varie de 7 à 13 et la longueur de la chaîne de bits n varie de 6 à 10.

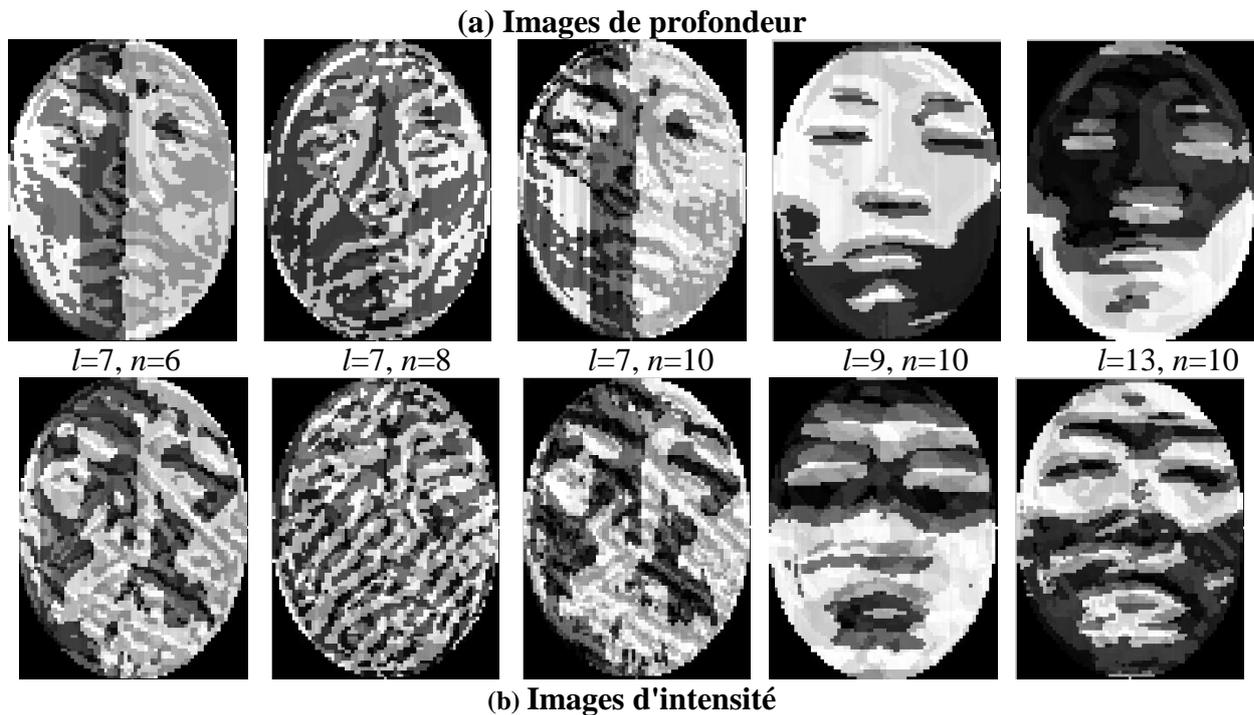


Figure 4.19: Le descripteur **BSIF** (Binarized Statistical Image)

4.3.3 Résultats

4.3.3.1 Base de données Bosphorus

La base de données Bosphorus [54, 334] est destinée à la recherche du traitement sur les tâches du visage humain 2D et 3D, y compris la reconnaissance de l'expression, la détection de l'action du visage, la reconnaissance des visages dans des conditions défavorables, la modélisation de visage déformable et la reconstruction du visage 3D. La base contient 4666 visages de 105 humains. Cette base est unique en trois aspects:

- répertoire riche d'expressions: Jusqu'à 35 expressions par sujet,
- poses de tête systématiques (13 rotations),
- variétés d'occlusions des visages (barbe et moustache, cheveux et lunettes).

Dans notre travail, nous utilisons 9 images pour chaque sujet. Nous avons utilisé un protocole d'évaluation pour la séparation des personnes en deux classes, client et imposteur. Le groupe de clients contient 90 sujets, tandis que le groupe imposteurs est divisé en 10 imposteurs pour l'ensemble d'évaluation et 5 pour l'ensemble de test. La distribution de photos de différents ensembles est présentée dans le tableau 4.8.

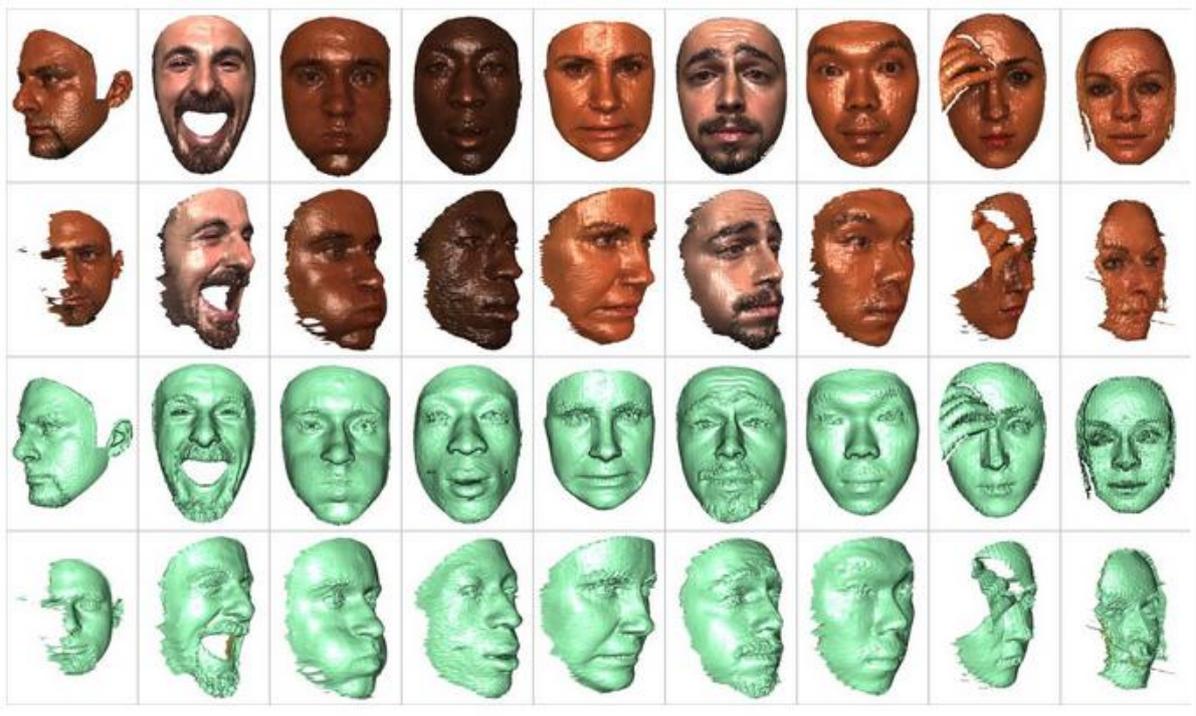


Figure 4.20: Echantillons avec et sans texture (Base de données du Bosphorus).

Tableau 4.8: Répartition des photos dans différents ensembles (base de données Bosphorus).

Ensemble	Client	Imposteur
Apprentissage	270 images (N_N, LFAU_27, UFAU_43)	0 images
Evaluation	270 images (E_HAPPY, LFAU_12, LFAU_28)	90 images (1:9)
Test	270 images (LFAU_34, UFAU_2, UFAU_2)	45 images (1:9)

4.3.3.2 Multi-échelle LBP (MLBP)

Pour cette partie, nous utilisons la méthode locale LBP sur la deuxième base de données Bosphorus. Les tableaux 4.3 et 4.9 montrent le taux d'erreur dans l'ensemble d'évaluation et de test de cette méthode d'extraction de caractéristiques pour les bases de données CASIA visage 3D et Bosphorus. Le nombre de points voisins (P) varie de 8 à 24 points et la valeur du rayon (R) varie de un à quatre pixels.

Dans le tableau 4.3, la méthode LBP pour l'information 3D donne plus de performances que l'information 2D pour les quatre valeurs du rayon (R). La fusion des quatre valeurs du rayon (MLBP) améliore les performances des informations 2D et 3D et la fusion multi-capteurs (2D + 3D) avec un EER = 0,95 % dans l'ensemble d'évaluation et TV = 94,54 % dans l'ensemble de test sur la base de données CASIA, ce qui n'est pas le cas de la base de données Bosphorus. La meilleure performance est obtenue dans le cas (P, R) = (24,4) avec et TEE = 4,92 % et TV= 91,38 %. Ceci peut être expliqué par la nature des bases de données, la variation d'acquisition etc...

Nous constatons que les performances d'authentification de visage s'améliorent avec l'augmentation du nombre de pixels du voisinage et du rayon considéré quel que soit la base de données d'application.

Tableau 4.9: Performance de la méthode d'extraction de caractéristiques MLBP dans l'ensemble d'évaluation et de test (base de données Bosphorus).

(P,R)	Image de profondeur (2D)					Image d'intensité (3D)					Image de profondeur et d'intensité (2D et 3D)			
	éval	test				éval	test				éval	test		
		TEE	TFA	TFR	TVNC		TEE	TFA	TFR	TV		NC	TEE	TFA
1- (8,1)	21.32	30.02	26.66	43.30	50	14.71	15.23	8.51	76.24	40	11.80	15.67	11.48	72.83
2- (16,2)	7.43	13.87	9.25	76.86	90	14.57	9.23	7.4	83.35	30	5.59	6.91	7.77	85.30
3- (24,3)	8.23	11.6	8.88	79.50	60	12.54	6.61	8.88	84.49	60	5.53	3.97	4.81	91.20
4- (24,4)	6.77	9.13	80.15	82.34	80	13.30	9.13	6.66	84.19	90	4.92	4.17	4.44	91.38
1+2+3+4 (MSLBP)	7.53	5.43	20.00	74.56	/	8.82	6.12	8.88	84.98	/	5.00	4.24	5.18	90.56

Tableau 4.10: Performance de la méthode d'extraction de caractéristiques CSL dans l'ensemble d'évaluation et de test (Base de données Bosphorus).

(P,R)	Image de profondeur (2D)					Image d'intensité (3D)					Image de profondeur et d'intensité (2D et 3D)				
	éval	test				éval	test				éval	test			
		TEE	TFR	TFA	TV		NC	TEE	TFR	TFA		TV	NC	TEE	TFR
mean	1- (8,1)	2.33	2.96	4.07	92.96	10	12.26	10.56	3.33	86.09	90	2.20	2.91	1.85	95.23
	2- (16,2)	2.14	2.69	4.44	92.86	30	11.33	11.60	2.59	85.80	40	1.73	3.30	1.11	95.58
	3- (24,3)	1.93	2.51	4.07	93.40	30	10.83	10.37	2.96	86.66	50	2.09	2.46	1.11	96.41
	4- (24,4)	2.89	8.34	2.22	89.43	50	12.01	10.83	3.33	85.82	90	2.88	5.08	1.11	93.80
	1+2+3+4	2.10	2.46	2.96	94.56	/	10.83	10.74	2.96	86.29	/	2.12	3.62	1.11	95.25
std	1- (8,1)	4.19	5.8	3.33	90.86	60	16.67	19.60	17.03	63.35	60	3.75	4.51	3.33	92.14
	2- (16,2)	1.54	1.43	1.48	97.08	30	11.85	11.23	10.37	78.39	20	2.17	2.86	1.11	96.02
	3- (24,3)	2.92	1.7	1.48	96.81	60	10	9.01	7.77	83.20	50	1.61	2.29	0.37	97.33
	4- (24,4)	2.31	2.41	2.22	95.35	30	8.18	8.02	6.29	85.67	30	1.50	1.85	1.11	97.03
	1+2+3+4	1.64	1.65	1.85	96.49	/	8.67	8.56	6.29	85.13	/	1.94	2.32	0.74	96.93
var	1- (8,1)	4.39	3.4	5.92	90.66	40	20.33	23.35	23.7	52.93	30	3.88	2.88	4.44	92.66
	2- (16,2)	1.98	1.53	5.18	93.28	1.98	16.40	20.32	15.92	63.75	60	1.62	1.85	2.59	95.55
	3- (24,3)	2.15	1.18	2.96	95.85	50	14.32	12.29	13.33	74.37	80	1.97	1.43	1.11	97.45
	4- (24,4)	2.2	1.58	1.48	96.93	40	12.62	10.83	13.7	75.45	20	2.19	1.50	1.11	97.38
	1+2+3+4	1.45	0.76	2.59	96.64	/	13.08	11.48	12.59	75.92	/	1.16	0.61	2.59	96.79
skewness	1- (8,1)	8.11	10.88	8.51	80.59	70	14.43	13.92	5.55	80.51	40	6.94	9.01	4.44	86.54
	2- (16,2)	6.05	13.60	4.81	81.58	60	15.14	16.96	4.44	78.59	90	5.38	11.11	2.22	86.66
	3- (24,3)	6.13	16.37	5.55	78.07	60	14.32	13.75	2.22	84.02	50	5.12	10.98	1.11	87.90
	4- (24,4)	6.16	13.67	3.7	82.61	30	15.71	17.08	2.59	80.3	60	5.80	13.03	0.74	86.22
	1+2+3+4	5.35	12.88	3.70	83.40	/	14.60	14.54	2.59	82.86	/	3.47	7.40	5.18	87.40
kurtosis	1- (8,1)	15.2	18.41	18.51	63.06	50	22.09	26.37	20.37	53.25	10	15.08	18.37	17.77	63.85
	2- (16,2)	10.25	13.58	8.14	78.27	60	28.05	30.19	29.62	40.1	10	10.32	13.35	8.51	78.12
	3- (24,3)	9.96	11.82	12.96	75.20	50	31.38	34.04	32.59	33.3	10	9.69	11.45	12.96	75.58
	4- (24,4)	15.96	16.34	16.66	66.98	10	28.93	31.01	31.85	37.13	20	16.46	17.13	12.96	69.90
	1+2+3+4	8.62	11.38	9.62	78.98	/	23.58	26.51	24.81	48.66	/	8.17	10.24	9.62	80.12
mean(24,3)+std(24,4)+var(24,3)+skewness(24,4)+kurtosis(24,4)	1.74	1.72	1.48	96.79	/	6.59	5.18	4.07	90.74	/	1.35	1.48	0.74	97.77	
mean(24,3)+std(24,3)+var(24,3)+skewness(24,4)	0.79	1.18	1.85	96.96	/	6.56	4.71	4.07	91.20	/	0.98	0.98	1.11	97.90	

4.3.3.3 Caractéristiques statistiques locales (CSL)

Pour cette partie, nous utilisons notre descripteur local (CSL). Le tableau 4.4 (base de données CASIA, visages 3D) et le tableau 4.10 (base de données Bosphorus) montrent le taux d'erreur dans l'ensemble d'évaluation et de test de la méthode CSL. Le nombre de points voisins (P) varie de 8 à 24 points et la valeur du rayon (R) varie de 1 à 4 pixels.

- Premièrement, la fusion des quatre rayons ($R = 1, 2, 3, 4$) pour les différents voisinages n'améliore pas les performances de cinq descripteurs statistiques disponibles.
- Nous remarquons aussi que, pour $R = 3, R = 4$ et le nombre de points $P = 24$ (taille maximale dans notre application), on obtient un meilleur résultat avec tous les descripteurs statistiques. Par

conséquent, l'augmentation du nombre de points dans le voisinage améliore la performance dans le cas de descripteurs statistiques.

- Les quatre descripteurs (*moyenne, écart-type, variance, skewness*) donnent à peu près les mêmes résultats. *Kurtosis* est le moins bon descripteur.
- La fusion des cinq paramètres de nos caractéristiques locales améliore l'authentification du visage. La performance sans *Kurtosis* est meilleure que la fusion des cinq paramètres statistiques.
- Nous avons obtenu de bonnes performance et stabilité avec TV= 96,43 % dans la base de données CASIA et TV = 97,90 % dans la base de données Bosphorus.

4.3.3.4 Extraction de caractéristiques par le descripteur BSIF (Binarized Statistical Image Features)

Toujours dans le but d'améliorer les performances de notre authentification de visage, nous nous intéressons à une technique statistique BSIF pour répondre à une question que nous nous sommes posée: les méthodes statistiques analysent-elles mieux l'image du visage ? L'étude et l'application de la méthode BSIF dans les deux bases de données nous donnent les résultats illustrés dans les figures 4.21 et 4.22.

- **Application à la base de données CASIA visage 3D**

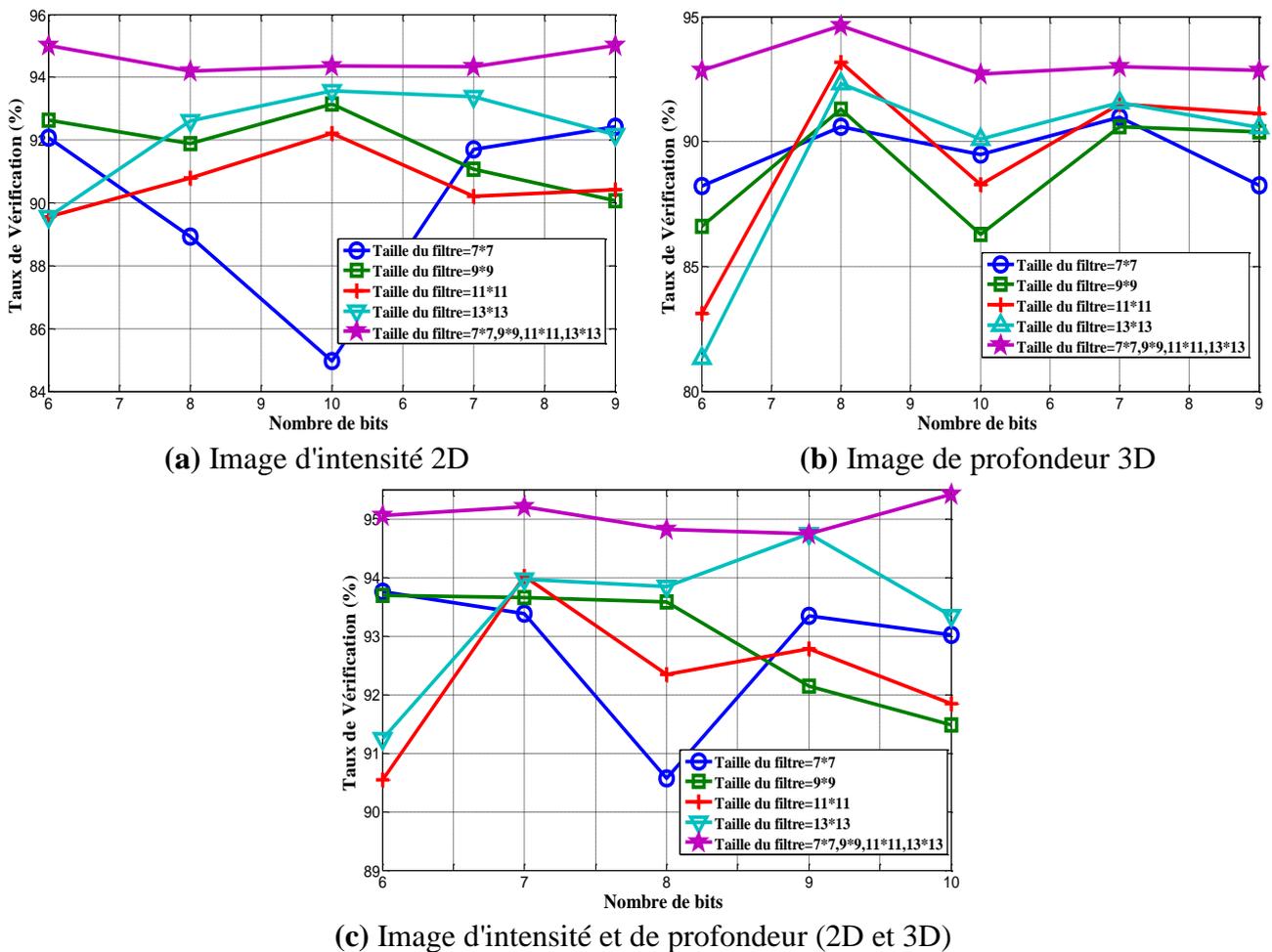


Figure 4.21: Performance de la méthode d'extraction de caractéristiques BSIF (base de données CASIA visage 3D).

Nous pouvons dire que:

- La fusion des filtres et la fusion multimodale (profondeur et intensité) donnent la meilleure performance.
- La taille du filtre affecte les performances de la reconnaissance du visage. Le plus grand des filtre est le plus important.
- Nous notons que la taille du filtre 13×13 donne le meilleur résultat pour les données 2D, 3D et (2D+3D).
- La meilleure performance obtenue par la fusion de quatre tailles de filtres et pour le nombre de bits égal à 10 pour un système multi-capteurs d'authentification de visage avec TEE = 1,9 % et TV = 95,42 %.

- Application à la base de données Bosphorus

L'application de la méthode BSIF sur la deuxième base de données utilisée dans notre travail donne les résultats ci dessous:

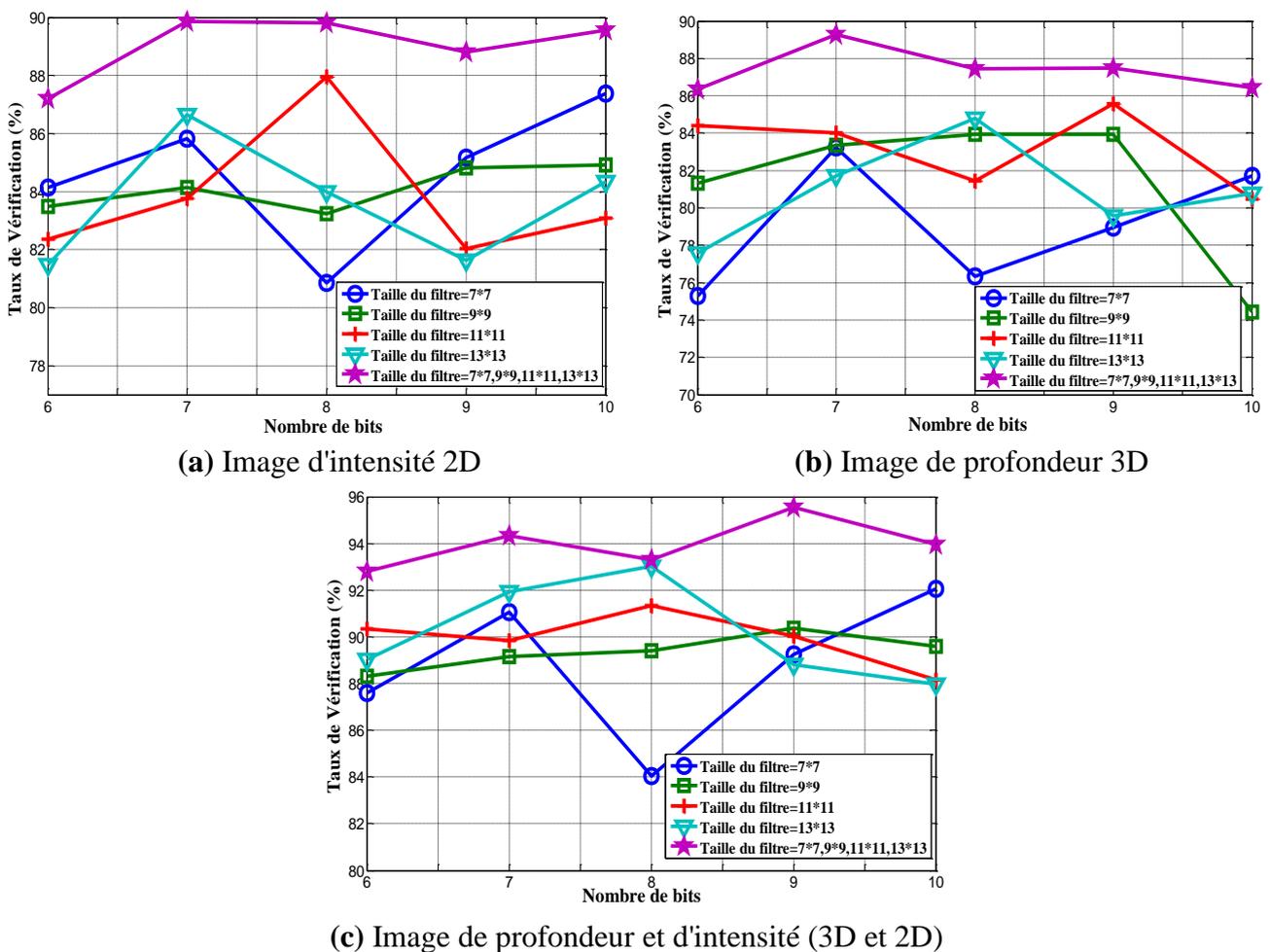


Figure 4.22: Performance de la méthode d'extraction de caractéristiques BSIF (base de données Bosphorus).

Dans le cas de la base 3D Bosphorus:

- La fusion des différents filtres et des deux modalités (profondeur et intensité) donne les meilleures performances.
- Les résultats sont équivalents pour les deux images profondeur (3D) et intensité (2D).

Chapitre 4: Vérification de visage 2D et 3D par CSL et les descripteurs locaux

- Le filtre 11×11 est meilleur que le filtre 13×13 pour les modalités profondeur et intensité. Ce n'est pas le cas pour la fusion des modalités où le filtre 13×13 reste le meilleur.

Nous pouvons dire à ce stade du travail que :

- La fusion des filtres est retenue ainsi que la fusion des modalités de profondeur et d'intensité.
- Le descripteur BSIF est performant pour des tailles du filtre supérieure à 9.

4.3.3.5 Extraction de caractéristiques par le descripteur LPQ (Local Phase Quantization)

La quatrième méthode d'extraction de caractéristiques locales étudiées dans la deuxième partie des résultats de ce chapitre est la LPQ. Les expériences sont menées sur les deux bases de données et nous donnent les résultats présentés dans les tableaux 4.11 et 4.12.

Tableau 4.11: Performances de la méthode d'extraction de caractéristiques LPQ dans l'ensemble d'évaluation et de test (base de données CASIA visage 3D).

$R=(M-1)/2$	Image de profondeur (2D)					Image d'intensité (3D)					Image de profondeur et d'intensité (2D et 3D)			
	éval	test				éval	test				éval	test		
	TEE	TFA	TFR	TV	NC	TEE	TFA	TFR	TV	NC	TEE	TFA	TFR	TV
1- 1	23.00	22.3	22.60	55.09	90	19.04	18.92	16.2	64.87	20	18.24	17.51	15.80	66.68
2- 2	14.40	14.41	13.8	71.78	100	11.15	10.86	7.6	81.54	70	10.35	9.66	6.60	83.73
3-3	7.77	8.44	8.8	82.76	90	8.38	7.44	6.6	85.96	40	5.77	5.38	5.80	88.82
4- 4	9.83	11.08	10.8	78.12	60	6.62	6.52	6	87.48	60	5.83	5.93	6.00	88.06
5- 5	10.44	13.5	11	75.50	70	4.82	5.02	5	89.97	80	5.374	5.64	4.80	89.56
6- 6	8.76	10.28	9.6	80.12	100	4.77	4.16	4.4	91.44	60	3.18	2.78	6.00	91.21
7- 7	7.63	8.62	9.8	81.58	90	4.35	3.65	4.2	92.14	40	3.64	2.44	4.60	92.96
8- 8	7.18	7.34	8.4	84.25	100	3.63	2.37	4.8	92.82	70	3.956	2.97	5.20	91.82
1:8	4.74	4.96	7.20	87.83	/	3.12	2.78	4.00	93.21	/	2.95	2.46	4.00	93.53
1:4	6.36	7.42	6.80	85.77	/	6.36	5.63	5.20	89.16	/	4.56	4.02	4.800	91.17
5:8	3.22	7.48	7.80	84.71	/	3.22	2.45	4.20	93.34	/	3.70	2.79	3.40	93.80

Tableau 4.12: Performances de la méthode d'extraction de caractéristiques LPQ dans l'ensemble d'évaluation et de test (base de données Bosphorus).

$R=(M-1)/2$	Image de profondeur (2D)					Image d'intensité (3D)					Image de profondeur et d'intensité (2D et 3D)			
	éval	test				éval	test				éval	test		
	TEE	TFA	TFR	TV	NC	TEE	TFA	TFR	TV	NC	TEE	TFA	TFR	TV
1- 1	32.20	35.45	32.96	31.58	30	27.68	36.79	25.18	38.02	20	24.81	31.45	25.18	43.35
2- 2	10.43	15.08	16.29	68.61	80	21.55	31.50	16.29	52.19	40	9.70	13.95	10.37	75.67
3-3	9.96	13.55	10.37	76.07	80	15.22	25.85	10.74	63.40	40	8.27	12.71	5.18	82.09
4- 4	8.33	7.58	9.25	83.16	70	15.03	23.25	7.77	68.96	90	6.24	7.35	5.18	87.45
5- 5	9.59	11.28	7.77	80.93	80	15.04	20.91	6.29	72.79	50	6.32	9.43	2.59	87.97
6- 6	7.3	9.06	6.66	84.27	70	11.7	19.18	5.55	75.25	70	4.10	7.95	3.33	88.71
7- 7	6.22	5.48	4.07	90.44	80	9.19	14.81	4.07	81.11	60	3.48	4.71	2.96	92.32
8- 8	7.03	6.83	5.55	87.60	70	8.58	21.40	3.33	75.25	80	4.38	9.11	1.85	89.03
1:8	4.13	4.56	5.18	90.24	/	7.42	15.33	2.59	82.07	/	2.86	4.00	3.33	92.66
1:4	4.11	5.13	11.85	83.01	/	11.51	20.27	9.25	70.46	/	4.24	6.39	4.81	88.79
5:8	5.16	4.51	4.44	91.03	/	7.00	15.13	2.59	82.27	/	3.72	2.85	3.40	93.74

Nous pouvons dire que:

- Les résultats obtenus sur les deux bases de données montrent l'importance de la fusion des modalités de profondeur et d'intensité.
- LPQ donne de mauvais résultats pour R inférieur à 4.
- La meilleure performance est obtenue pour la fusion des quatre grands R .

Chapitre 4: Vérification de visage 2D et 3D par CSL et les descripteurs locaux

- Les performances d'authentification de visage sont équivalentes pour les deux bases de données.
- Les résultats sont encourageants mais pas suffisants. Notons que les taux de reconnaissance obtenus sont stables d'une base de données à une autre.

4.6.3.6 Fusion des méthodes d'extraction de caractéristiques locales

Dans cette partie du travail nous tentons de faire la fusion des différents descripteurs et de les appliquer sur les deux bases de données.

Les simulations menées nous donnent les résultats présents dans les tableaux 4.13 et 4.14.

Tableau 4.13: Performances de quatre descripteurs et leur fusion dans l'ensemble d'évaluation et test (base de données CASIA visage 3D).

Extraction de caractéristiques	Image de profondeur (2D)				Image d'intensité (3D)				Image de profondeur et d'intensité (2D et 3D)			
	éval	test			éval	test			éval	test		
		TEE	TFA	TFR		TV	TEE	TFA		TFR	TV	TEE
1- LBP	4.62	4.98	4.40	90.61	5.36	4.63	3.80	91.56	0.95	1.25	4.20	94.54
2-CSL	5.04	2.03	2.60	95.36	3.24	1.87	3.60	94.52	1.20	0.88	2.80	96.32
3-BSIF	1.91	2.20	2.80	95.00	4.30	2.55	4.60	92.84	1.90	1.78	2.80	95.42
4-LPQ	3.22	7.48	7.80	84.71	3.22	2.45	4.20	93.34	3.70	2.79	3.40	93.80
1+2	1.93	1.86	2.60	95.53	3.25	2.35	3.60	94.04	1.97	1.04	2.20	96.75
2+3	1.00	0.87	2.80	96.32	3.81	2.47	4.00	93.52	1.91	1.20	2.60	96.20
2+4	2.51	1.69	2.80	95.50	2.71	1.64	3.40	94.95	1.85	1.48	2.20	96.32
1+2+3	0.77	1.01	3.00	95.98	3.22	1.95	3.80	94.24	1.71	0.99	2.20	96.80
1+2+4	1.62	1.34	3.00	95.65	2.16	0.98	3.40	95.62	1.55	0.76	2.60	96.64
2+3+4	0.47	0.92	3.40	95.67	2.55	1.47	3.60	94.92	1.50	1.01	2.20	96.78
1+2+3+4	1.65	1.25	3.00	95.74	2.20	0.94	3.60	95.46	1.34	0.67	2.60	96.72
1+3	0.70	1.62	3.40	94.98	3.55	2.17	4.00	93.82	1.90	1.92	2.80	95.28
1+4	3.23	4.30	4.20	91.50	2.26	0.96	3.60	95.44	2.22	1.95	2.60	95.44
3+4	2.14	2.24	3.20	94.55	2.65	1.47	3.80	94.72	2.15	1.88	2.60	95.52
1+3+4	1.95	2.25	3.20	94.54	2.48	1.42	3.20	95.37	1.96	1.54	2.80	95.65

Tableau 4.14: Performances de quatre descripteurs et leur fusion dans l'ensemble d'évaluation et test (base de données Bosphorus).

Les résultats obtenus sur la base de données CASIA sont encourageants notamment dans le cas de la fusion des descripteurs LBP+CSL+BSIF avec un TEE=1.71 % et un TV= 96.80 %. Notons aussi que le meilleur descripteur reste la CSL avec un TEE=1.20 % et un TV=96.32 %.

D'après les résultats obtenus dans le tableau 4.14, on peut dire que notre méthode proposée CSL est en tête suivie de la BSIF puis de LPQ. LBP tient la dernière place dans la reconnaissance faciale dans les deux cas des images de profondeur et d'intensité.

La fusion des trois descripteurs LBP+CSL+BSIF donne les meilleures performances avec un TV=98.41 % et un TEE=1.20 %, ce qui est très satisfaisant dans le cas de la base de données Bosphorus et TV=96.80 % et TEE=1.71 % sur la base de données CASIA 3D.

Enfin retenons de tous ces résultats que:

- la CSL que nous proposons est meilleure que les quatre descripteurs étudiés.
- la fusion des deux modalités (profondeur et intensité) ainsi que la fusion des trois descripteurs (CSL+BSIF+LPQ) donnent de meilleurs résultats que ceux obtenus antérieurement.

Extraction de caractéristiques	Image de profondeur (2D)				Image d'intensité (3D)				Image de profondeur et d'intensité (3D et 2D)			
	éval		test		éval		test		éval		test	
	TEE	TFA	TFR	TV	TEE	TFA	TFR	TV	TEE	TFA	TFR	TV
1- LBP	6.77	9.13	80.15	82.34	13.30	9.13	6.66	84.19	4.92	4.17	4.44	91.38
2-CSL	0.79	1.18	1.85	96.96	6.56	4.71	4.07	91.20	0.98	0.98	1.11	97.90
3-BSIF	3.38	7.11	4.07	88.81	6.96	8.41	4.07	87.50	2.33	3.70	0.74	95.55
4-LPQ	5.16	4.51	4.44	91.03	7.00	15.13	2.59	82.27	3.72	2.85	3.40	93.74
1+2	1.53	2.04	1.11	96.83	4.80	4.54	3.70	91.75	0.90	1.30	0.74	97.95
2+3	1.49	1.20	1.11	97.67	4.85	4.34	1.85	93.8	1.19	1.28	0.37	98.34
2+4	0.93	1.33	1.11	97.55	3.60	6.04	2.96	2.96	0.72	1.03	1.11	97.85
1+2+3	1.40	1.25	1.11	97.62	4.35	4.34	2.22	93.43	1.35	1.20	0.37	98.41
1+2+4	1.03	1.01	1.48	97.50	3.50	5.97	2.96	91.06	1.08	1.11	0.74	98.14
2+3+4	0.58	01.16	1.11	97.72	4.16	6.17	1.48	92.34	0.65	1.20	1.11	97.67
1+2+3+4	0.56	97.72	1.11	97.72	3.82	5.55	1.85	92.59	1.35	1.43	0.37	98.19
1+3	4.20	8.49	4.07	87.43	5.679	6.64	2.96	90.39	2.85	4.00	1.11	94.88
1+4	5.80	5.65	4.81	89.53	7.04	12.88	2.59	84.51	2.16	4.56	2.59	92.83
3+4	0.81	2.49	5.92	91.58	4.48	6.88	3.33	89.77	1.72	2.49	2.22	95.28
1+3+4	3.32	6.14	3.70	90.14	4.08	6.44	3.70	89.85	1.77	2.69	1.11	96.19

Tableau 4.15: Comparaison avec d'autres méthodes locales mentionnées dans la littérature.

Date	Auteur	Modalité	Descripteur	Base de données	Meilleurs résultats
2010	Wang et al [335]	3D	Fusion de LBP, Gabor, PCA	CASIA	TV=91.71%
2010	Y.A. Li et al [336]	3D	Geometric Features	CASIA	TR=91.1% (rank-1)
2011	Jin et al [337]	3D	2D Gabor et LBP	FRGC v2	TV=82.8% à TFA=0.1%
2013	Tang et al [225]	3D et 2D	self-adaptive voting LBP (V-LBP)	BJUT-3D FRGC v2	TR=92.4% (rank-1) TR=94.89% (rank-1)
2013	Berretti et al [338]	3D	Histogram of orientations	BU-3DFE UF-3D	TR=87.5% (rank-1) Frontal TV=96.2% LeftTR= 54.7 %% (rank-1)
			geometric histogram	BU-3DFE UF-3D	TR=84% (rank-1) Frontal TV= 94.3 % LeftTR= 58.5% (rank-1)
2013	Smeets et al [339]	3D	Mesh SIFT	Bosphorus	TR=97.73% (rank-1)
2013	Zhao et al [340]	3D	Bayesian Belief Network (BBN)	Bosphorus	TV=94.2%
2014	Ming [341]	3D	Orthogonal Spectral Regression (ROSR)	Bosphorus CASIA	TV=93.95% TR=96.13%
2014	Notre travail	3D et 2D	CSL	CASIA	TV=96.32%
			CSL	Bosphorus	TV=97.90%
			LBP+CSL+BSIF	CASIA	TV=96.80%
			LBP+CSL+BSIF	Bosphorus	TV=98.41%

4.3.3.7 Comparaison

Le tableau 4.15 présente une comparaison de nos résultats avec ceux de la littérature de reconnaissance de visage en 3D et 2D.

Le tableau montre que:

- Notre système d'authentification de visage multimodal proposé donne de bonnes performances comme indiqué dans le tableau 4.15.
- Les résultats pour les deux bases de données donnent presque les mêmes performances avec un TV = 97,90 % pour la base de données CASIA et TV = 98,41 % pour la base de données Bosphorus. Ceci montre la qualité de notre système proposé.

4.4. Conclusion

Dans ce chapitre, nous avons présenté un algorithme automatique d'authentification de visage par la fusion de deux systèmes multimodaux (multi-algorithmes: la fusion de plusieurs descripteurs et multi-capteurs: la fusion de l'image de profondeur (3D) et d'intensité (2D)). Tout d'abord, nous avons corrigé la rotation de la tête par l'algorithme ICP (Iterative Closest Point), puis présenté notre algorithme de prétraitement. Pour l'extraction de caractéristiques, six méthodes locales ont été utilisées et qui sont:

- Descripteur MLBP (Multi-échelle Local Binary Patterns).
- Caractéristiques Statistiques Locales proposées (CSL).
- Ondelettes de *Gabor*.
- Descripteur SIFT (Scale-Invariant Feature Transform).
- Descripteur LPQ (Local Phase Quantization).
- Descripteur BSIF (Binarized Statistical Image Features).

La classification des caractéristiques, par la métrique cosinus après réduction d'espace de données par analyse en composantes principales (ACP), suivie du modèle discriminant linéaire amélioré de *Fisher* (EFM). La fusion au niveau des scores est effectuée par un bon classificateur à deux classes appelées Support Vector Machines (SVM).

L'application est réalisée sur les bases de données CASIA visage 3D et le Bosphorus. Selon notre protocole d'évaluation proposé pour traiter des problèmes majeurs dans le domaine de la reconnaissance faciale 3D et multimodal en tenant compte des variations d'éclairage, d'expression, de petite rotation de la tête et des changements combinés dans l'expression sous éclairage. Nos contributions sont les suivantes:

1. Le développement d'un algorithme de prétraitement automatique.
2. Une nouvelle méthode d'extraction de caractéristiques (Caractéristiques Statistiques Locales (CSL)) est proposée. Elle est basée sur le calcul de paramètres statistiques dans un voisinage du pixel tels que la moyenne, l'écart type, la variance, etc...
3. Étude de la fusion de deux systèmes multimodaux (Multi-algorithmes et Multi-capteurs).
4. Étude de plusieurs méthodes d'extraction de caractéristiques pour mieux comprendre leur complémentarité.
5. Application pour la première fois d'un nouveau descripteur BSIF (Binarized Statistical Image Features) sur les images de profondeur (3D).

Dans ce chapitre on a deux parties de résultats. En première partie, nous avons utilisé quatre descripteurs (MLBP, CSL, ondelettes de *Gabor* et SIFT) et les résultats obtenus sont appliqués sur une seule base de données (CASIA visages 3D). Dans cette partie, nous pouvons dire que:

- PCA + EFM donne une mauvaise performance pour l'information de profondeur (3D). Nous avons obtenu: TV=89.36 % pour l'information de profondeur (3D), TV=92.26 % pour l'information d'intensité (2D) et TV=93.14 % pour la fusion multimodale (3D + 2D).
- La méthode LBP pour l'information 3D donne plus de performances que l'information 2D pour les quatre valeurs du rayon (R). La fusion des quatre valeurs du rayon (MLBP) améliore les performances des informations 2D et 3D, et la fusion multi-capteurs (2D + 3D) avec un EER = 0,95 % et TV = 94,54 %.
- La fusion des quatre rayons ($R = 1, 2, 3, 4$) pour les différents nombres de points voisins n'améliore pas les performances de cinq descripteurs statistiques CSL. Nous remarquons aussi que, pour $R = 3, R = 4$ et le nombre de points $P = 24$ (taille maximale dans notre application),

on obtient un meilleur résultat pour tous les descripteurs statistiques. Par conséquent, l'augmentation du nombre de points dans le voisinage améliore la performance dans le cas de descripteurs statistiques. Les quatre descripteurs (*moyenne, écart-type, variance, skewness*) donnent à peu près les mêmes résultats. *Kurtosis* est le moins bon descripteur.

- La fusion des cinq paramètres de nos caractéristiques locales améliore l'authentification du visage. La performance sans *Kurtosis* est meilleure que la fusion des cinq paramètres statistiques avec TEE= 1,20 % et TV = 96,32 % pour la fusion multimodale (2D et 3D).
- Les meilleurs résultats par les ondelettes de *Gabor* sont obtenus lorsque la résolution $\lambda= 4$ (TEE = 1,57 % et TV = 96,13 %). La fusion de cinq résolutions n'améliore pas la performance de vérification de visages.
- La fusion de quatre LBP ($R=1, 2, 3, 4$ et $P=8, 16, 24$) plus SIFT donne le meilleur résultat avec un TEE = 2,48 % et TV=94. 73 %.
- La fusion de notre descripteur CSL avec le descripteur MLBP + SIFT donne les meilleurs résultats avec TEE = 0,98 % et TV = 97,22 %.

Dans la deuxième partie des résultats nous avons utilisé quatre descripteurs (MLBP, CSL, LPQ et BSIF). L'application est réalisée sur deux bases de données CASIA visages 3D et Bosphorus. Dans cette partie, nous pouvons dire que:

- Les résultats obtenus sur les deux bases de données montrent l'importance de la fusion des modalités profondeur et intensité.
- La fusion des quatre valeurs de rayon (MLBP) améliore les performances avec un EER = 0,95 % et TV = 94,54 % sur la base de données CASIA, ce qui n'est pas le cas de la base de données Bosphorus. La meilleure performance est obtenue dans le cas $(P, R) = (24,4)$ avec et TEE = 4,92 % et TV= 91,38 %.
- Nous avons obtenu de bonnes performance et stabilité avec TV= 96,43 % dans la base de données CASIA et TV = 97,90 % dans la base de données Bosphorus pour le descripteur (CSL).
- Le descripteur BSIF est performant pour des tailles du filtre supérieur à 9. La fusion des filtres est retenue pour les deux bases de données.
- LPQ donne des mauvais résultats pour R inférieur à 4 et la meilleure performance est obtenue pour la fusion des quatre grands R . Les performances d'authentification de visage sont équivalentes pour les deux bases de données.
- La fusion des trois descripteurs LBP+CSL+BSIF donne les meilleures performances avec un TV=98.41 % et un TEE=1.20 % ce qui est très satisfaisant dans le cas de la base de données Bosphorus, et TEE=1.71 % et TV=96.80 % pour la base de données CASIA 3D.
- Le tableau 4.15 présente une comparaison de notre travail avec des travaux de la littérature de reconnaissance de visage en 3D et 2D. Notre système d'authentification du visage multimodal proposé donne de bonnes performances.

Pour les travaux futurs, nous proposons de:

- ✓ améliorer la détection du nez (depuis notre algorithme qui utilise seulement le point le plus saillant),
- ✓ améliorer la phase de classification (PCA+EFM),
- ✓ exploiter les histogrammes des descripteurs locaux utilisés.
- ✓ utiliser la meilleure base de données 3D existante (Face Recognition Grand Challenge FRGC v2).

Nous envisagerons toutes ces perspectives dans le chapitre 5.

5.1 Introduction

En raison de son pouvoir discriminant et la simplicité de calcul, le descripteur LBP (Local Binary Pattern) est devenu une approche très populaire dans diverses applications de vision par ordinateur. Cela est vrai pour, par exemple, la détection et l'analyse des visages, la biométrie, l'analyse d'images médicales, le mouvement et l'analyse de l'activité et la récupération des bases de données d'images ou vidéo [342, 329]. Plusieurs chercheurs en reconnaissance de visage sont devenus intéressés par LBP. Les auteurs dans leur papier IEEE PAMI en 2006 [76] ont montré son grand succès dans la reconnaissance des visages. Selon [329], la LBP n'est pas considérée seulement comme un opérateur de texture simple, mais constitue le fondement d'une nouvelle direction de recherche importante pour les descripteurs binaires locaux sur l'image et la vidéo. Ces dernières années, différentes variantes de LBP ont été proposées pour améliorer sa robustesse et accroître son pouvoir discriminant et son applicabilité à différents types de problèmes [76] (voir section 2.2.2.2).

Motivés par le succès et l'utilisation généralisée de la LBP et ses variantes dans la reconnaissance faciale 2D, nous vous proposons ce chapitre comme une extension importante de notre travail récent [351], où un cadre général de reconnaissance faciale 2D est basée sur la fusion multi-échelles multi-descripteurs locaux binaires. Le résumé de ce travail est présenté dans la section 5.2.

5.2 Résumé de notre travail ICIP'2014 [343]

La plupart de l'état de l'art des systèmes de vérification de visage divise l'image de visage en régions plus petites et la similitude est mesurée région par région. La décision finale est basée sur la fusion de ces scores. En faisant cela, plus d'informations de localisation peuvent être préservées. Chaque région correspond à une zone locale de l'image et est de dimension relativement faible, ce qui signifie moins de temps à calculer. En outre, par rapport à la représentation globale, cette méthode basée sur la région est beaucoup plus robuste à l'éclairage et à la variation d'expression. La raison en est que la variation d'illumination et d'expression à l'intérieur de l'ensemble de l'image est plus grande que celle à l'intérieur de chaque région.

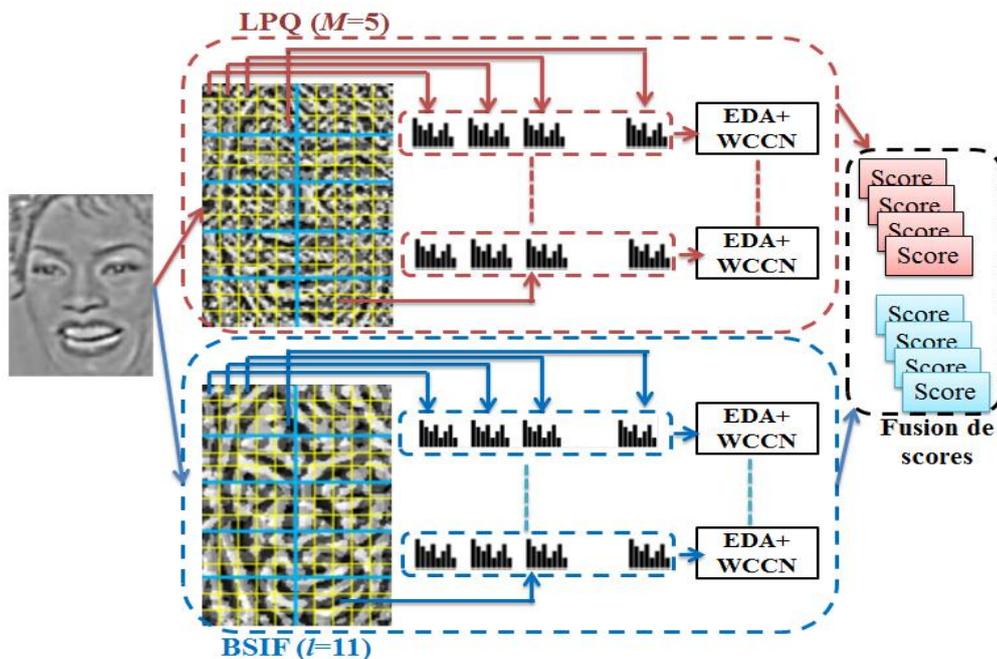


Figure 5.1: Schéma général de notre proposition ICIP'14 [343].

Dans ce travail, après l'application du descripteur binaire local, l'image est divisée en régions (10 régions, 5 lignes et 2 colonnes) et chaque région est divisée en 15 blocs (3 lignes et 5 colonnes). Pour

chaque bloc et pour toutes les échelles, les statistiques de LBP (Local Binary Pattern), LPQ (Local Phase Quantization) et BSIF (Binarized Statistical Image Features) sont résumées par des histogrammes. Les histogrammes des différents blocs appartenant à la même région sont concaténés pour former un vecteur de caractéristiques représentant un descripteur spécifique à une échelle spécifique. Cette stratégie de prendre l'histogramme de l'ensemble de la région plutôt que les petits blocs, comme fait dans [344], rend le système plus robuste à l'occlusion et se traduit également par un faible nombre de scores. Cependant, la concaténation des 15 histogrammes des blocs provoque toujours des vecteurs de grande dimension contenant des informations redondantes. Pour cela, nous proposons l'utilisation de l'EDA (Exponential Discriminant Analysis) pour surmonter le problème de petite taille de l'échantillon (small sample size problem, SSSP) sans écarter l'information discriminante contenue dans l'espace nul de la LDA. La base de projection EDA est suivie par WCCN (Within Class Covariance Normalization) pour déclasser l'effet de la direction de haute inter variabilité et améliorer la discrimination de la similarité cosinus pour chaque région. Enfin, les scores pour chaque région sont fusionnés par SVM. La figure 5.1 donne un schéma de principe général du système proposé.

Afin d'assurer une reproductibilité des expériences et la comparaison avec d'autres méthodes, nous avons testé notre approche sur les bien connues bases de données 2D XM2VTS et LFW en utilisant les protocoles communs.

5.2.1 Expérience dans la base de données XM2VTS

Nous avons effectué des expériences de vérification du visage sur l'ensemble de la configuration LPI. Des détails sur la base de données XM2VTS peuvent être trouvés dans la section 3.2.5.1. Les images ont été coupées à une taille standard de 190×160 . Les taux globaux d'erreur dans l'ensemble d'évaluation et de test à l'aide des descripteurs de visage individuels, plusieurs descripteurs et un état de l'art sont présentés dans le tableau 5.1.

Le tableau 5.1 montre que les taux de vérification obtenus par MLPQ à l'aide de quatre échelles et MBSIF avec 5 échelles (fenêtres) surpassent tout les résultats de l'état de l'art selon notre connaissance. En outre, en utilisant de multiples descripteurs, les résultats de la fusion connaissent une diminution importante pour le taux global d'erreurs dans l'ensemble d'évaluation et de test. Les multiples descripteurs (MLBP+MLPQ+MBSIF), suivie par (EDA+WCCN) ont atteint un taux d'erreur totale dans l'ensemble de test égale à 0,31 %, ce qui est loin des meilleurs résultats rapportés sur la base de données XM2VTS (0,96 %).

Tableau 5.1:Taux d'erreur total sur l'ensemble d'évaluation et de test (base de données XM2VTS, Protocole de Lausanne I).

	Descripteur	Evaluation $2 \times \text{TEE}$ (en %)	Test TFR+TFA (en %)
Descripteur simple	MLBP ^{u2} ($R=2+4+6+8$, $P=8$)	0.98	2.41
	MLPQ ($M=5+9+13+17$)	0.27	0.81
	MBSIF ($l=7+9+11+13+15$)	0.28	0.72
	LBPH_MAP [345]	/	2.84
	LBPHMM [346]	/	2.74
	Gabor : ICB2006-Best [347]	1.63	0.96
Descripteurs multiples	MLBPH^{u2}+MLPQH+MBSIFH	0.11	0.31
	Notre approche (chapitre 3) : Gabor+composante S (HSV) [318]	1	2.66

Taux d'erreur total dans l'ensemble d'évaluation = $2 \times \text{TEE}$ (TEE: taux d'égale erreur).

Taux d'erreur totale dans l'ensemble de test=TFR+TFA (TFR: Le taux de faux rejet et TFA: le taux de fausse acceptation).

5.2.2 Expérience dans la base de données LFW

La base de données « Labeled Faces in the Wild (LFW) » [348] est constituée d'images du visage collectées sur le site d'information Yahoo ! News. Des exemples d'images sont présentés dans la figure 5.2. Aucune contrainte sur les paramètres de prise de vue n'a donc été imposée. La base LFW contient 13233 images de 5479 personnes différentes.



Figure 5.2: Exemples d'images de la base de données LFW.

Parmi ces personnes, 1680 sont représentées par au moins deux images.

- Les protocoles d'expérimentation

En plus des images, les concepteurs de la base ont mis au point un protocole expérimental permettant de définir précisément les mesures de performance. A cette fin, deux vues sont proposées. Elles prennent la forme de listes de paires d'images avec à chaque fois autant de paires d'images positives (les deux images représentent la même personne) que de paires d'images négatives (les deux images représentent des personnes différentes). Les figures 5.3 et 5.4 présentent des exemples des mêmes et différentes personnes respectivement. La «vue 1» est utilisée pour la mise au point des algorithmes et la sélection du modèle. Elle se compose d'un jeu de données d'entraînement de 2200 paires et d'un jeu de test de 1000 paires. La «vue 2» n'est utilisée que pour le rapport final de performance et contient 10 séries de 600 paires, soit 6000 paires en tout. Ces 10 séries constituent les 10 étapes de la validation croisée utilisée pour la mesure de performance, laquelle est donnée par le taux moyen de classification exacte et l'erreur standard sur cette moyenne.



Figure 5.3: Exemples d'images de la même personne de la base de données LFW.



Figure 5.4: Exemples des mêmes et différentes paires de visages dans la base de données LFW.

Dans cette expérience, nous utilisons les images alignées (LFW-a) [349]. Les images ont été coupées à une taille standard de 130×90. Elles sont photo métriquement normalisées par l'approche PS (preprocessing sequence approach). Dans la base de données LFW, la plupart des sujets ont un seul échantillon. Nous prenons des sujets ayant plus de deux images pour la formation EDA et WCCN. Pour la fusion de scores, nous utilisons 5 séries pour la formation de SVM et 4 séries pour les tests de de l'SVM et la dernière série pour l'évaluation des performances. Une comparaison du taux moyen de classification des systèmes proposés sur les données de test «vue 2» avec les résultats des études de l'état de l'art est présentée dans le tableau 5.2.

La première remarque du tableau 5.2 est l'efficacité éprouvée de la MBSIF qui est légèrement meilleure que la MLPQ laquelle est légèrement meilleure que la MLBP^{u2}. Il est également intéressant de mentionner que notre système basé sur multi-échelle LPQ (MLPQ +EDA+WCCN) surpasse le système présenté dans [344] en utilisant le même descripteur (91.27±0.71 % contre 89.00±1.55 %). Notre système est plus résistant aux défauts d'alignement parce que nos régions représentent les plus grands histogrammes que dans [344] et aussi en raison de l'efficacité d'EDA+ WCCN. En comparant le cadre proposé pour les méthodes existantes sur l'état de l'art dans la même configuration «Unrestricted» en utilisant plusieurs descripteurs, le taux moyen de classification de notre système (MLBP + MLPQ + MSIFT + EDA + WCCN + similarité cosinus + la fusion par SVM) atteint 93,03 ± 0,82 % qui est l'un des meilleurs résultats trouvés dans la littérature.

Tableau 5.2: (Mean ± Standard Error Score) de nos systèmes sur la base de données LFW «Unrestricted protocol view 2».

	Descripteur	$\mu \pm S_E$
Descripteur simple	MLBP ^{u2} (R=2+4+6+8,P=8)	88.60±0.87
	MLPQ (M=5+9+13+17)	91.27±0.71
	MBSIF (7+9+11+13+15)	91.40±0.94
	MLPQ(EXT+INT)[344]	89.00±1.55
	CMD [355]	91.70±1.1
Descripteurs multiples	MLBP ^{u2} +MLPQ+MBSIF (proposé)	<u>93.03±0.82</u>
	LBP+TPLBP+FPLBP+SIFT [23]	89.5±0.51
	MLPQ+MLBP (EXT+INT) [344]	90.02 ±1.45
	face.com r2011b [350]	91.30±0.3
	CMD+SLBP [351]	92.58±1.36
	DM+PCA fusion [351]	92.05 ±0.45

5.3 Description de l'ensemble du système de vérification faciale 2D+3D proposé

Dans ce travail, nous présentons un système (figure 5.5) de vérification de visage multimodal assez efficace basé sur la fusion de l'information locale. Dans la première étape, une phase de prétraitement est nécessaire pour le remplissage des trous et le rééchantillonnage des données 3D et 2D. L'algorithme ICP (Iterative Closest Point) est utilisé pour aligner tous les visages 2D et 3D avec le

premier visage 3D (expression neutre) et l'approche PS (preprocessing sequence approach) est utilisée pour réduire l'influence de la variation de l'éclairage pour les images 2D. Les images faciales 3D et 2D sont divisées en 10 régions, chacune divisée en 15 petites blocs. Pour chacune des méthodes utilisées et pour chaque bloc, l'extraction de caractéristiques locales est résumée par les histogrammes correspondants (expression 5.1). Les histogrammes des différents blocs appartenant à la même région sont concaténés pour former un vecteur de caractéristiques représentant un descripteur spécifique à une échelle spécifique. Pour la réduction de dimensionnalité, nous proposons d'utiliser EDA (Exponential Discriminant Analysis) [139] couplée à WCCN (Within Class Covariance Normalization) [352] pour rétrograder l'effet des directions de haute inter variabilité et renforcer la discrimination pour chaque région. Enfin, les scores des différentes régions correspondant aux différents descripteurs à différentes échelles sont normalisés par la normalisation Z_{norm} (expression 5.2) [353] et sont fusionnés par simple somme.

$$h(i) = \sum_{x,y} B(Image_p(x,y) = i), \quad i \in [0, 2^P - 1], \quad B(v) \begin{cases} 1 & \text{lorsque } v \text{ est vrai} \\ 0 & \text{autrement} \end{cases} \quad (5.1)$$

Où $h(i)$ est un histogramme qui est invariant à la translation et à la rotation de l'image dans une certaine mesure. P est le nombre de bits.

$$S_n = \frac{S - \text{mean}(S)}{\text{std}(S)} \quad (5.2)$$

S est un vecteur de scores de similarité de taille N (où N est la taille de la galerie), $\text{mean}(S)$ and $\text{std}(S)$ sont la moyenne et l'écart type du vecteur S respectivement.

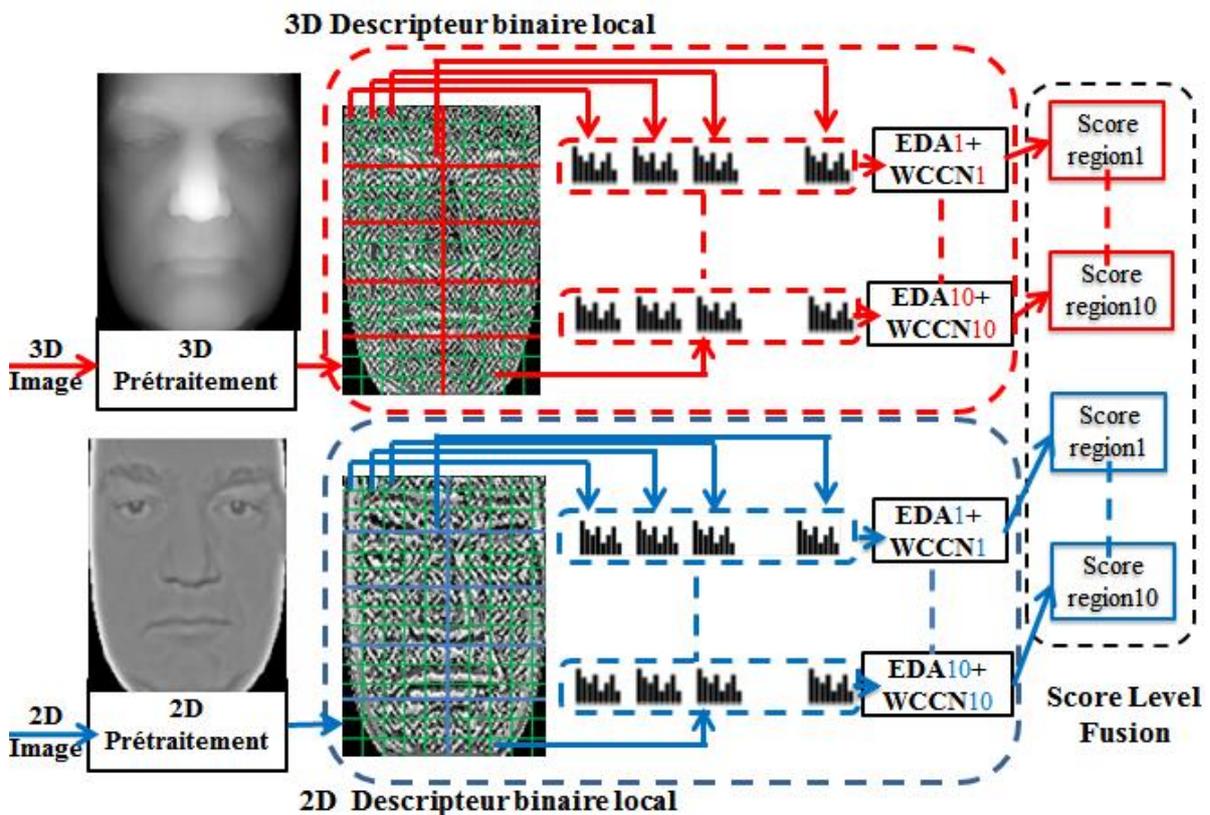


Figure 5.5: Schéma général de notre système proposé.

5.3.1 Prétraitement de visage 2D et 3D

Nous avons appliqué nos expériences sur la grande base de données FRGC v2 (Face Recognition Grand Challenge version v2) [55] pour la reconnaissance faciale 3D et nous avons utilisé encore une autre base de données (CASIA 3D) pour le test de la stabilité de nos expériences. En outre, les visages 3D sont brouillés et contiennent des points aberrants (figure 5.6 (a)). Les images des bases de données utilisées sont principalement acquises à partir du niveau d'épaule. Il est bien connu que l'emplacement du nez est autour du centre du visage. Nous détectons d'abord le bout du nez en utilisant l'approche proposée dans [354].

5.3.1.1 Détection du bout du nez

Dans un premier temps, l'information de courbure locale en chaque point, appelée indice de forme (shape index), est calculée. L'indice de forme est proposé par *Dorai* et *Jain* [355] et représente les informations de forme locale des points. L'indice de forme au point p est calculé en utilisant la valeur maximale ($k_{max}(p)$) et la valeur minimale ($k_{min}(p)$) de la courbure locale.

$$\text{indice de forme } (p) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{k_{max}(p)+k_{min}(p)}{k_{max}(p)-k_{min}(p)} \quad (5.3)$$

L'indice de forme est indépendant des coordonnées du système et peut s'adapter à la forme de la surface. La valeur zéro représente une calotte sphérique. Les points dont l'indice de forme est de l'ordre de 0,85 à 1 sont groupés pour former des zones possibles de la pointe du nez. Comme il peut y avoir de petites régions causées par les cheveux ou d'autres facteurs, les petites régions (<70 points) sont enlevées. Après cela, les régions possibles d'extrémité du nez sont étiquetées comme S_{Int} . Notons que ces régions comprennent les zones autour des oreilles, les yeux, les lèvres et le menton. Après cela, deux modèles de pointe du nez sont utilisés [356]. Tout d'abord, un modèle est utilisé pour faire correspondre le visage afin de déterminer quelles sont les régions dans S_{Int} qui ressemblent localement à une pointe du nez. Ensuite, pour les régions où le modèle de pointe du petit nez est sélectionné, nous sélectionnons un modèle de la plus grande pointe du nez est marqué pour sélectionner le réel de pointe du nez p_{nt} .

Une sphère de rayon r_l centrée à la pointe du nez p_{nt} est ensuite utilisée pour recadrer le visage 3D (figure 5.6 (b)) et son visage 2D correspondant (figure 5.7 (b)). Une valeur $r_l = 80 \text{ mm}$ a été choisie dans nos expériences. Une fois que le visage est coupé, les points aberrants sont éliminés. Un point aberrant d'observation est défini dans [255]. Les points aberrants sont ceux dont la valeur est supérieure à un seuil d_t de l'un de ses 8 voisins. d_t est calculée automatiquement à l'aide:

$$d_t = \mu + 0.6 \sigma \quad (5.6)$$

Où: μ est la valeur moyenne des points voisins et σ son écart-type.

La suppression des pointes peut entraîner des trous dans le visage 3D (figure 5.6 (c)), qui sont remplis en utilisant l'interpolation cubique pour combler les données manquantes. Ceci apparaît souvent dans les yeux et les zones de la bouche. Après avoir éliminé les pointes aberrantes, les visages sont acquis à des distances différentes du capteur 3D. Ils ont donc des résolutions différentes. Nous normalisons les images du visage 3D et 2D sur une grille carrée uniforme de 1 mm , afin de nous assurer que tous les visages ont la même résolution.

Après découpage, élimination des pointes aberrantes, remplissage des trous et normalisation, nous alignons tous les visages 3D et 2D avec le premier visage 3D (expression neutre) en utilisant l'algorithme ICP (Iterative Closest Point) [357] (section 4.2.1.1).

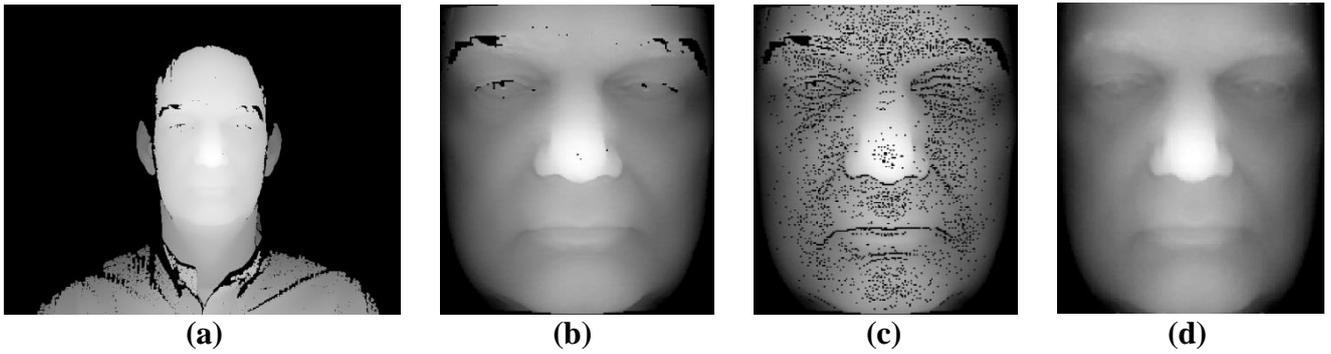


Figure 5.6: Prétraitement de l'image de profondeur: (a) données 3D, (b) image 3D coupée, (c) élimination des pointes aberrantes, (d) image de profondeur après étape de prétraitement.

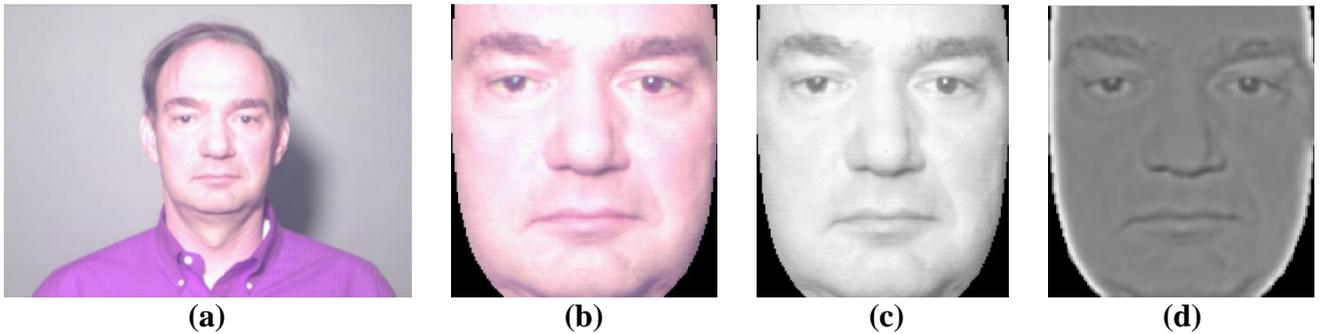


Figure 5.7: Prétraitement de l'image 2D: (a) données 2D, (b) image 2D coupée, (c) image d'intensité, (d) normalisation photométrique par l'approche PS (preprocessing sequence approach).

Suivant alignement par l'ICP, une sphère de rayon $r_2 = 75 \text{ mm}$ centrée sur le bout du nez est utilisée pour recadrer le visage 3D et 2D (figure 5.6.d et figure 5.7 (b)). Pour l'image couleur RVB (figure 5.7 (b)), nous avons utilisé l'image d'intensité (figure 5.7 (c)). Comme on le sait, la variation de l'éclairage influence fortement la présentation des images d'intensité (figure 5.8 (a)). L'approche PS [351] est utilisée pour réduire l'influence de la variation d'illumination. La figure 5.8 présente des images d'intensité de visage avant et après normalisation photométrique.

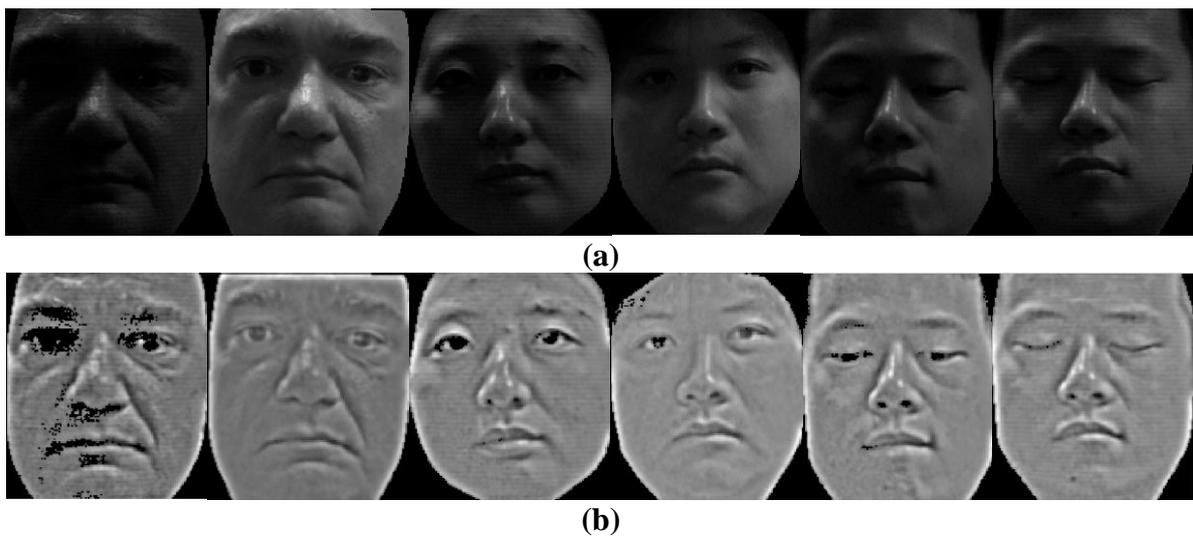


Figure 5.8: Images de visage 2D: (a) avant et (b) après normalisation photométrique par PS.

5.3.2 Extraction des caractéristiques

5.3.2.1 Descripteur *LBP* (Local Binary Patterns)

Nous utilisons dans ce chapitre le descripteur *LBP* pour l'extraction des caractéristiques des images de profondeur 3D et d'intensité 2D. Pour les images de profondeur nous utilisons l'opérateur *LBP* avec nombre des points P égale 10 (pour plus de détails sur le descripteur *LBP*, on peut consulter la section 4.2.2.1). Mais pour l'extraction de caractéristiques des images d'intensité, nous utilisons la *LBP* uniforme avec un nombre de points égal à 8.

- **LBP uniforme**

Un sous-ensemble de 2^P codes binaires, appelé modèles uniformes et définis dans [358], peut être utilisé pour représenter les spots, les fins de ligne, les bords et les coins. La mesure de l'uniformité, $U(x)$, présentée par l'expression 5.7, enregistre le nombre de transitions spatiales dans la configuration binaire et le modèle uniforme qui contient au plus deux transitions de bits, c'est-à-dire $U(x) \leq 2$. Le modèle uniforme contient au total $(P - 1) P + 2$ codes binaires. L'opérateur *LBP* uniforme ($LBP^{u2}_{P,R}$) est défini comme suit:

$$LBP^{u2}_{P,R}(x, y) = \begin{cases} I(LBP_{P,R}(x, y)) & \text{si } U(LBP_{P,R}) \leq 2, I(z) \in [0, (P - 1)P + 2] \\ (P - 1)P + 2 & \text{si } \text{non} \end{cases} \quad (5.7)$$

Où

$$U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^P |s(g_p - g_c) - s(g_{p-1} - g_c)| \quad (5.8)$$

$u2$ représenté dans l'expression 5.7 indique que la définition se rapporte à des modes uniformes avec une valeur $U \leq 2$. Si $U(x)$ est inférieur à 2, le pixel en cours est marqué par une fonction d'indexation $I(z)$. Sinon, il sera assigné à la valeur $(P - 1) P + 2$.

La fonction d'index, $I(z)$, contenant $(P - 1) P + 2$ indices, est utilisée pour attribuer un index particulier à chacune des configurations uniformes. Un exemple de codes uniformes pour $P=8$ est présent dans la figure 5.9.

L'utilisation d'un code *LBP* uniforme a deux avantages. Le premier est le gain en mémoire et en temps de calcul. Le deuxième est que LBP^{u2} permet de détecter uniquement les textures locales importantes, comme les spots, les fins de ligne, les bords et les coins (figure 4.5.b) pour des exemples de ces textures particulières. En effet, *Ojala et al.* ont montré que les *LBP*s uniformes contiennent plus de 90 % d'information d'une image.

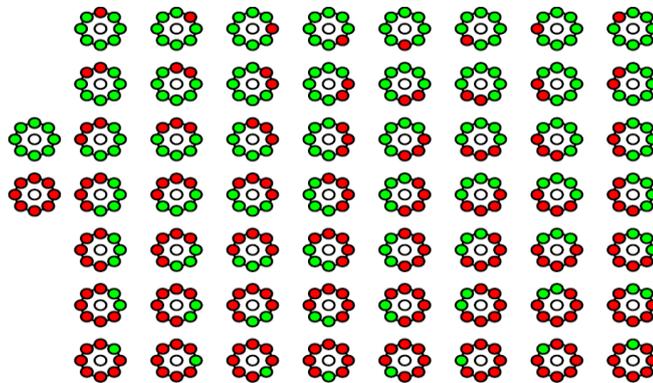


Figure 5.9: LBP uniforme $P=8$.

La figure 5.10 présente des exemples d'extraction de caractéristiques des images de profondeur et d'intensité par LBP.

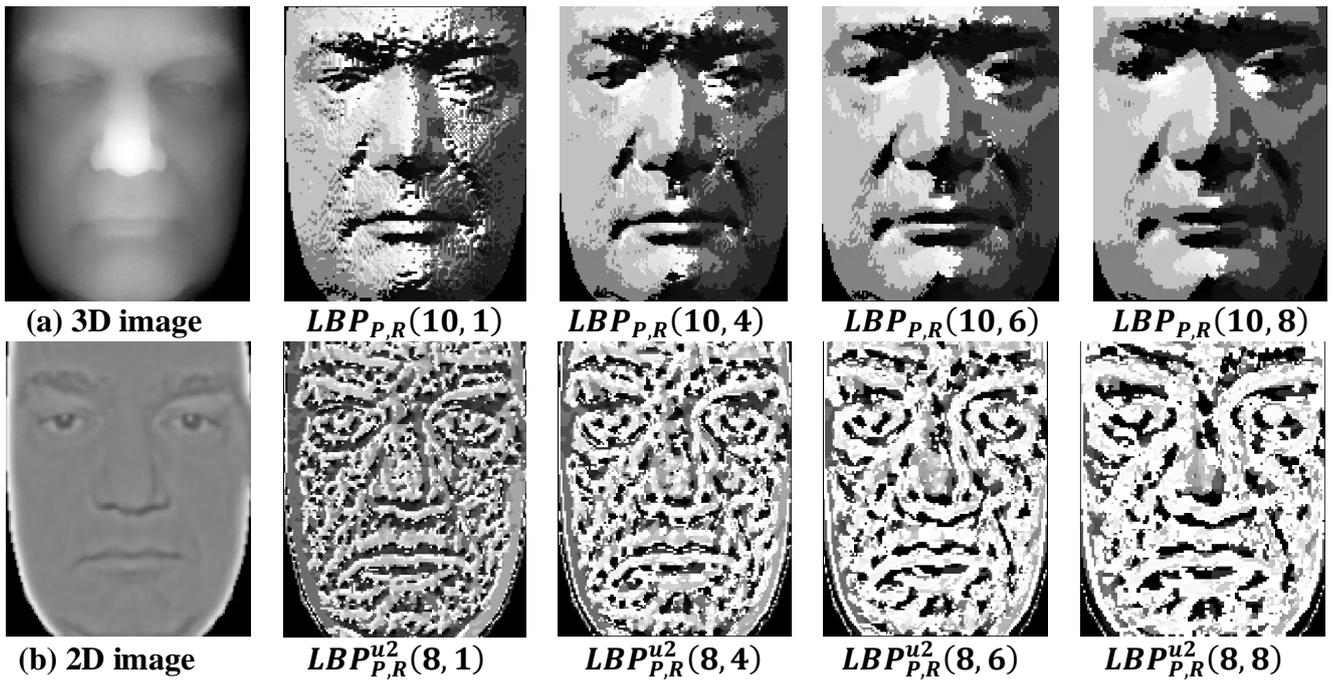


Figure 5.10: LBP (Local Binary Patterns) (a) image 3D, (b) image 2D.

5.3.2.2 Descripteur LPQ (Local Phase Quantization)

Nous utilisons la multi-échelle LPQ (MSLPQ) pour l'extraction de caractéristiques des images 3D et 2D (section 4.3.1). La représentation multi-échelle PLQ peut être obtenue en faisant varier la taille du filtre M . La figure 5.11 présente une illustration des images LPQ 3D et 2D avec différentes échelles dans lesquelles la taille de la fenêtre PLQ prend des valeurs différentes.

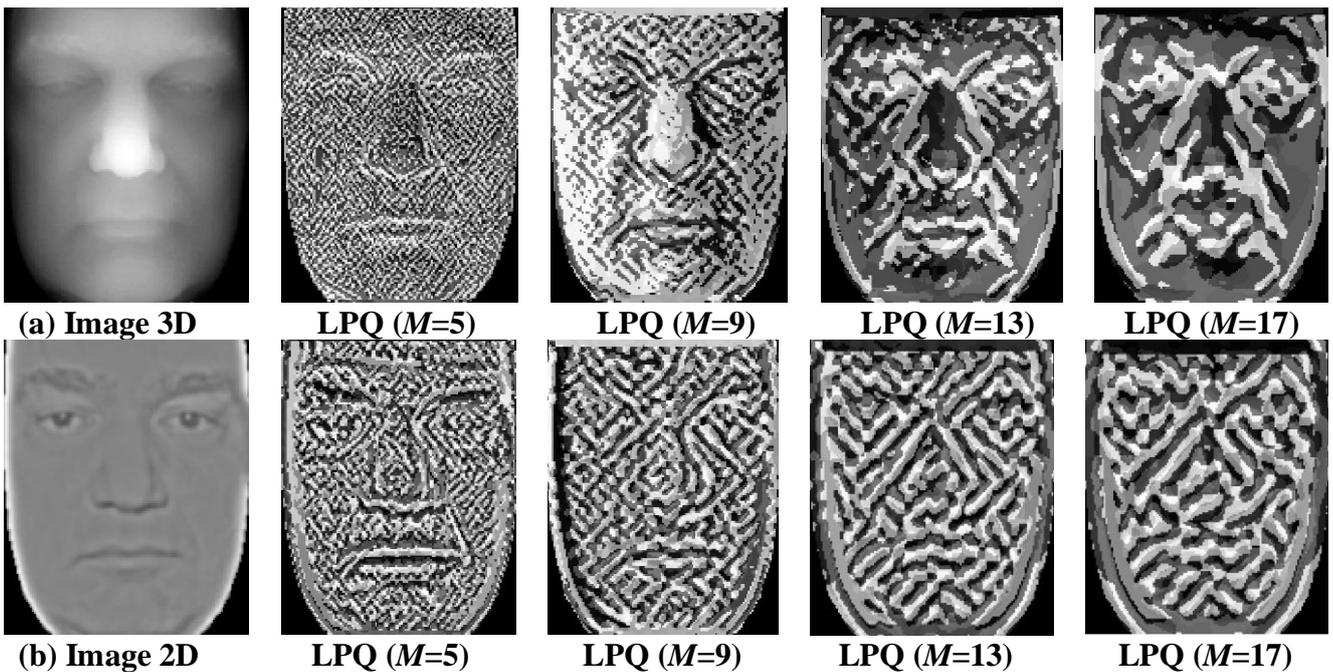


Figure 5.11: Local Phase Quantization (LPQ): (a) 3D image, (b) 2D image

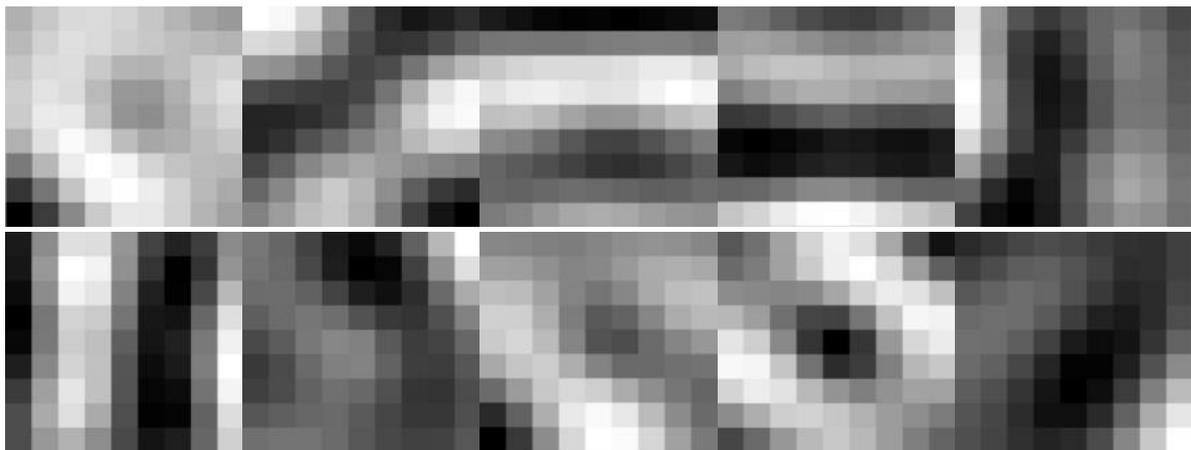
5.3.2.3 Descripteur BSIF (Binarized Statistical Image Features)

Le descripteur BSIF est proposé récemment par *Kannala et al.* [154] pour l'extraction de caractéristiques des images de textures et de visage 2D. Nous avons présenté en détails ce descripteur dans la section 4.3.2. Nous avons utilisé le descripteur original BSIF pour l'extraction des caractéristiques des images de profondeur 3D et d'intensité 2D dans le chapitre 4. L'ensemble des filtres utilisés par le descripteur BSIF original [154] sont construits en se basant sur 13 images naturelles qui sont présentées sur la figure 4.17. Dans ce chapitre, nous proposons l'utilisation des images d'apprentissage 3D et 2D prétraitées pour la construction de ces filtres.

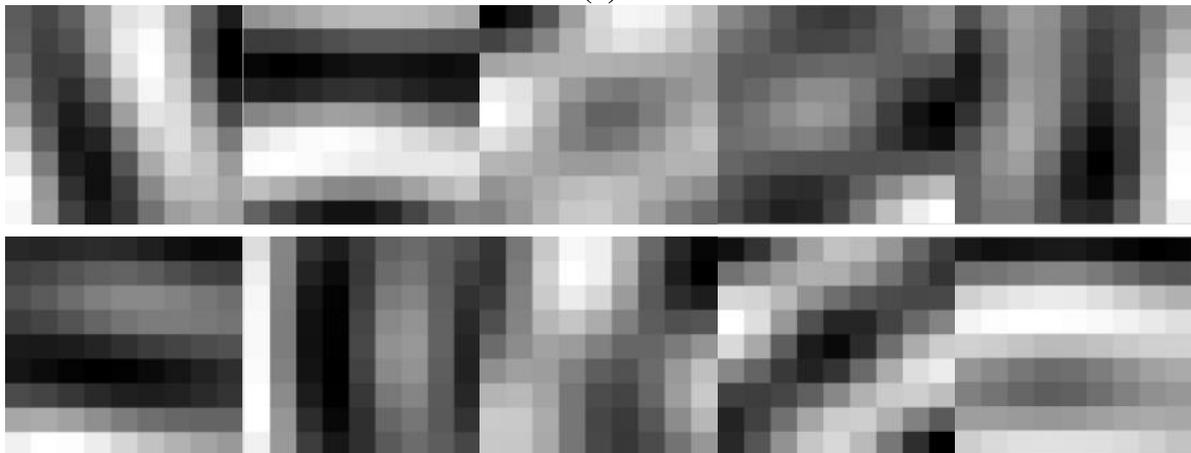
Les images prétraités 3D et 2D dans la phase d'apprentissage sont d'abord faites de moyenne nulle (la valeur moyenne de chaque image est soustraite) et sont divisées par leur écart-type. Nous sélectionnons au hasard 50000 blocs pour la construction des filtres (section 4.3.2.1 présente les étapes de construction des filtres).

La figure 5.12 montre les filtres obtenus avec la taille ($l = 9$) et le nombre de bits ($n = 10$) avec l'utilisation des images prétraitées 3D et 2D de la base de données FRGC v2. A notre connaissance, c'est le premier travail qui utilise les images de visage 3D et 2D pour la construction des filtres utilisés par le descripteur BSIF. Le code BSIF correspondant est présenté sur la figure 5.13.

- BSIFo: BSIF original (les filtres W_i sont construits par 13 images naturelles).
- BSIFn: BSIF nouveau (les filtres W_i sont construits par images 3D et 2D prétraitées du visage) où la taille du filtre l varie de 7 à 13.



(a)



(b)

Figure 5.12: Filtres construits de taille 9×9 , (a) pour image 3D et (b) pour image 2D

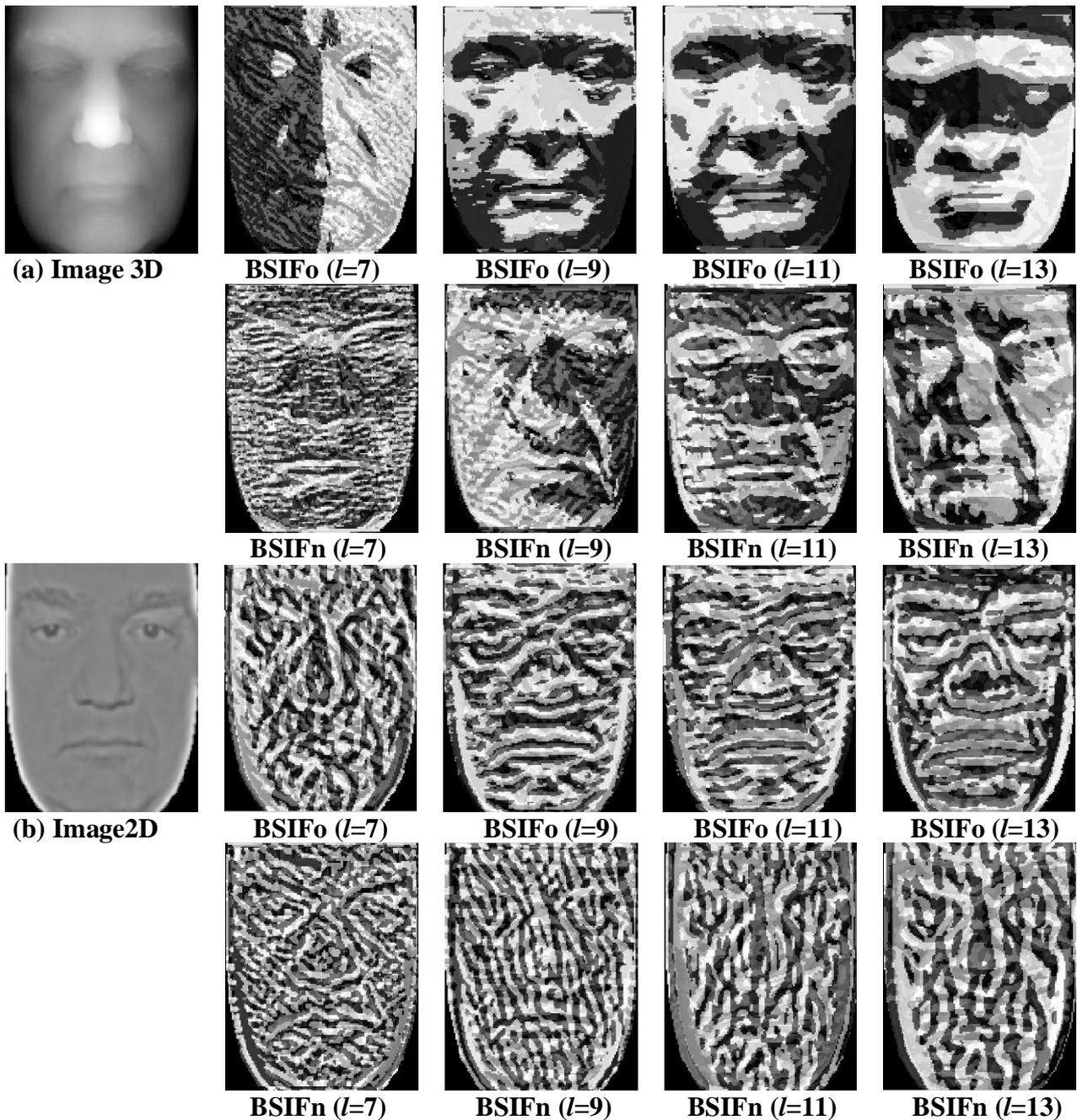


Figure 5.13: Descripteur BSIF (Binarized Statistical Image Features) (a) image 3D, (b) image 2D

5.3.2.4 Descripteur Statistical LBP proposé

Notre descripteur proposé est inspiré du descripteur LBP. L'opérateur LBP calcule un code pour chaque pixel par seuillage de sa valeur avec celle de son voisin et convertit ce code en un nombre décimal. L'objectif de notre méthode est l'augmentation de la précision et la création d'un nouvel espace de caractéristiques locales caractérisées par une variation des paramètres statistiques, qui sont: la moyenne, la médiane, la variance, l'asymétrie et l'aplatissement.

Nous remplaçons donc un pixel i par le calcul de ses paramètres statistiques ($Statistical_{p_1, R_1}(i)$) avec son voisin P_1 de rayon R_1 . Enfin, nous calculons le code binaire ($Statistical LBP_{p_1, R_1, p_2, R_2}(i_c)$) pour le pixel i_c par seuillage de son code statistique ($Statistical_{p_1, R_1}(i_c)$) avec les codes statistiques voisins ($Statistical_{p_1, R_1}(i_{p_2})$) (expression 5.9).

Le code Statistical LBP d'un pixel i_c est donné par:

$$\text{Statistical LBP}_{P_1, R_1, P_2, R_2}(i_c) = \sum_{p_2=0}^{P_2-1} s(\text{Statistical}_{p_1, R_1}(i_{p_2}) - \text{Statistical}_{p_1, R_1}(i_c)) 2^{p_2} \quad (5.9)$$

Où: i_c et i_{p_2} sont respectivement les valeurs du pixel central et P_2 sont les pixels voisins se trouvant à proximité du cercle de rayon R_2 (figure 5.14.). Sa fonction $s(x)$ est définie comme:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (5.10)$$

Statistical présente: la moyenne, la variance, la médiane, l'asymétrie et l'aplatissement, qui sont définis comme suit:

$$\text{mean}_{P_1, R_1}(i) = \frac{1}{P_1} \sum_{p_1=0}^{P_1-1} i_{p_1} \quad (5.11)$$

Où: i et i_{p_1} sont respectivement les valeurs du pixel central et P_1 les pixels voisins appartenant au cercle avec un rayon R_1 (figure 5.14).

La médiane d'un pixel i est la valeur numérique qui sépare la moitié supérieure des i_{p_1} pixels voisins, de la moitié inférieure.

La variance est l'un de plusieurs descripteurs d'une distribution de probabilité qui décrit dans quelle mesure les nombres se trouvent à partir de la moyenne (valeur attendue). Elle est définie par:

$$\text{Var}_{P_1, R_1}(i) = \frac{1}{P_1} \sum_{p_1=0}^{P_1-1} (i_{p_1} - \text{mean}_{P_1, R_1}(i))^2 \quad (5.12)$$

L'asymétrie (Skewness) d'une distribution ou d'un ensemble de données est positive si la queue de droite (à valeurs hautes) est plus longue ou grosse, et négative si la queue de gauche (à valeurs basses) est plus longue ou grosse. Elle est définie par:

$$\text{skewness}_{P_1, R_1}(i) = \frac{\frac{1}{P_1} \sum_{p_1=0}^{P_1-1} (i_{p_1} - \text{mean}_{P_1, R_1}(i))^3}{\left(\frac{1}{P_1} \sum_{p_1=0}^{P_1-1} (i_{p_1} - \text{mean}_{P_1, R_1}(i))^2 \right)^{3/2}} \quad (5.13)$$

L'aplatissement (Kurtosis) est une mesure qui indique si les données sont culminées ou plates par rapport à une distribution normale. Il est défini par:

$$\text{kurtosis}_{P_1, R_1}(i) = \frac{\frac{1}{P_1} \sum_{p_1=0}^{P_1-1} (i_{p_1} - \text{mean}_{P_1, R_1}(i))^4}{\left(\frac{1}{P_1} \sum_{p_1=0}^{P_1-1} (i_{p_1} - \text{mean}_{P_1, R_1}(i))^2 \right)^2} \quad (5.14)$$

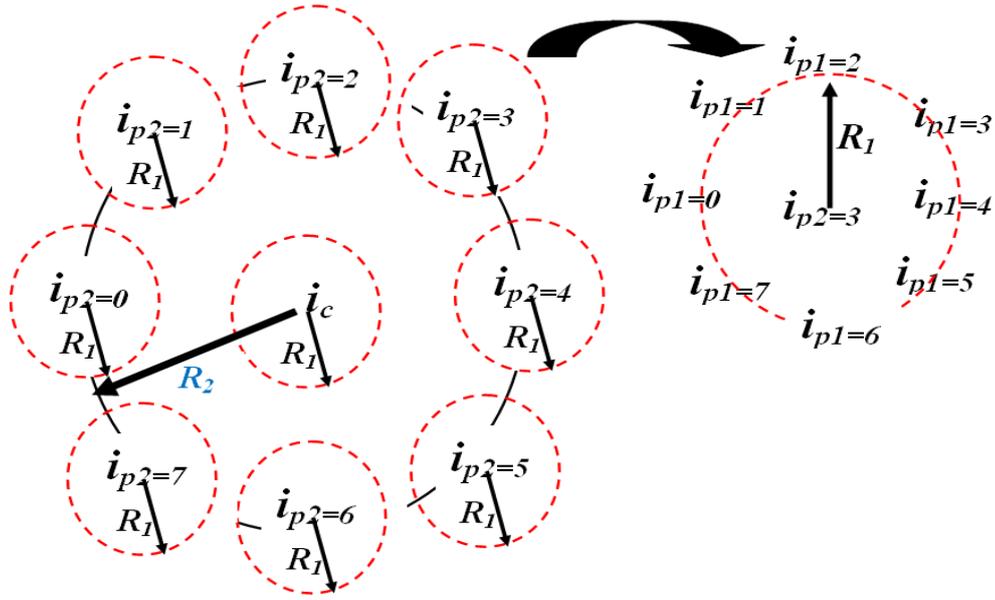


Figure 5.14: Descripteur Statistical LBP (Statistical Local Binary Patterns) $P_1=8, P_2=8$.

Les images Statistical LBP correspondant aux visages 3D et 2D sont présentées dans les figures 5.15 et 5.16 ($P_1 = 32, R_1 = 3, P_2 = 10, R_2 = 4$).

5.3.3 EDA (Exponential Discriminant Analysis) [73]

Soit un ensemble de données X représentées par N échantillons $\{x_1, x_2, \dots, x_N\}$ qui prennent des valeurs dans l'espace de dimension D , avec chaque échantillon appartenant à une classe $C, \{L_1, L_2, \dots, L_C\}$. En supposant que le nombre d'échantillons dans la classe L_i est représentée par N_i , on a $N = \sum_i N_i$.

Les matrices de dispersion totale S_t , inter-classes S_b et intra-classes S_w sont définies comme suit:

$$S_t = \sum_{i=1}^N (x_i - m)(x_i - m)^T = H_t H_t^T \quad (5.15)$$

$$S_b = \sum_{i=1}^C N_i (m_i - m)(m_i - m)^T = H_b H_b^T \quad (5.16)$$

$$S_w = \sum_{i=1}^C \sum_{x_j \in L_i} (x_j - m_i)(x_j - m_i)^T = H_w H_w^T \quad (5.17)$$

$$S_t = S_w + S_b \quad (5.18)$$

Où m_i représente la moyenne de la classe i , m la moyenne globale du jeu de données X .

Nous avons:

$$H_t = [(x_1 - m), (x_2 - m), \dots, (x_N - m)] \quad (5.19)$$

$$H_w = [A_1 - m_1 \cdot e_1^T, \dots, A_C - m_C \cdot e_C^T] \quad (5.20)$$

$$H_b = [\sqrt{N_1}(m_1 - m), \dots, \sqrt{N_C}(m_C - m)] \quad (5.21)$$

Où $e_i = (1, 1, \dots, 1)^T \in R^{N_i}$ et A_i est la matrice de données de la classe i .

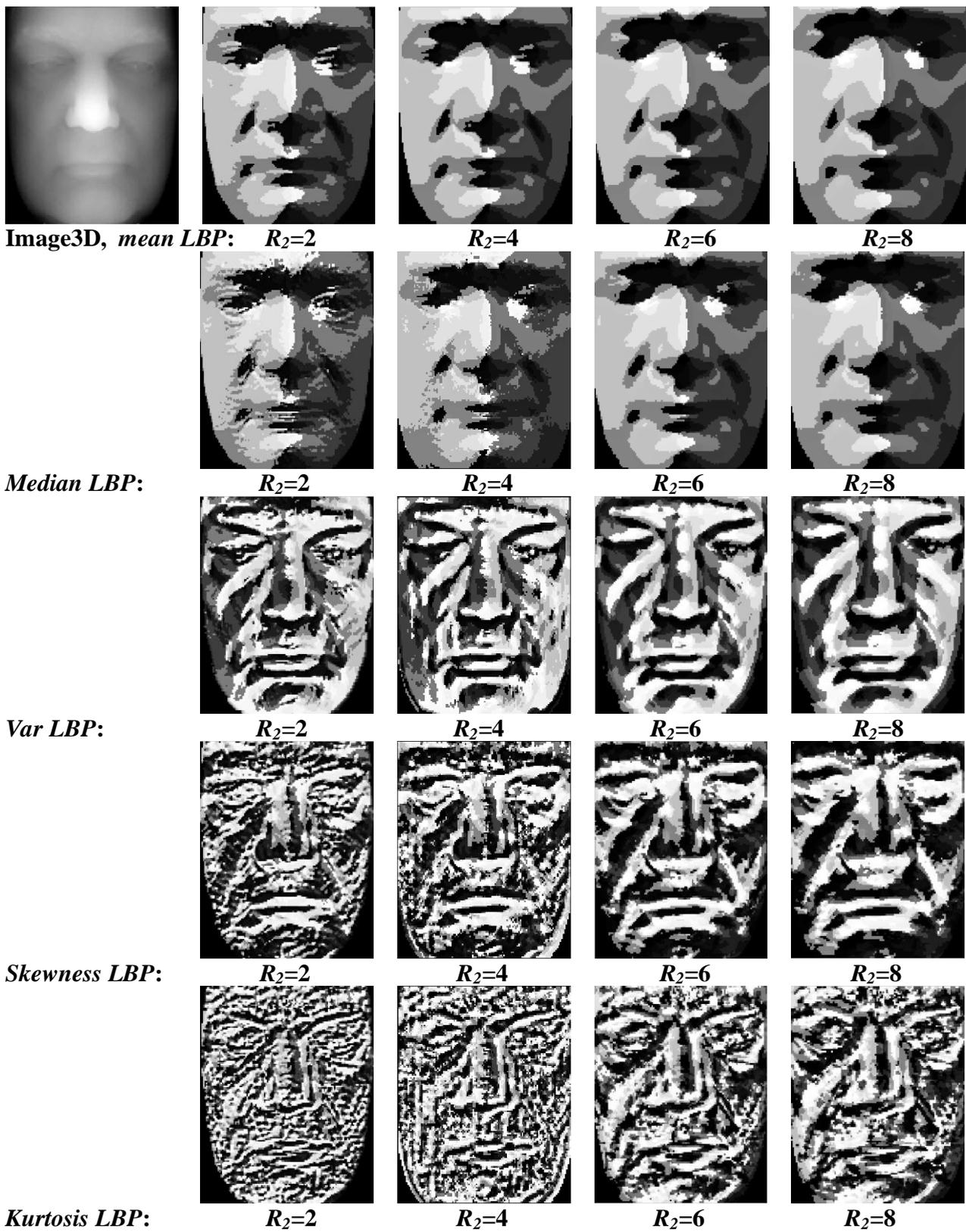


Figure 5.15: *Statistical LBP* pour les images de visage 3D.

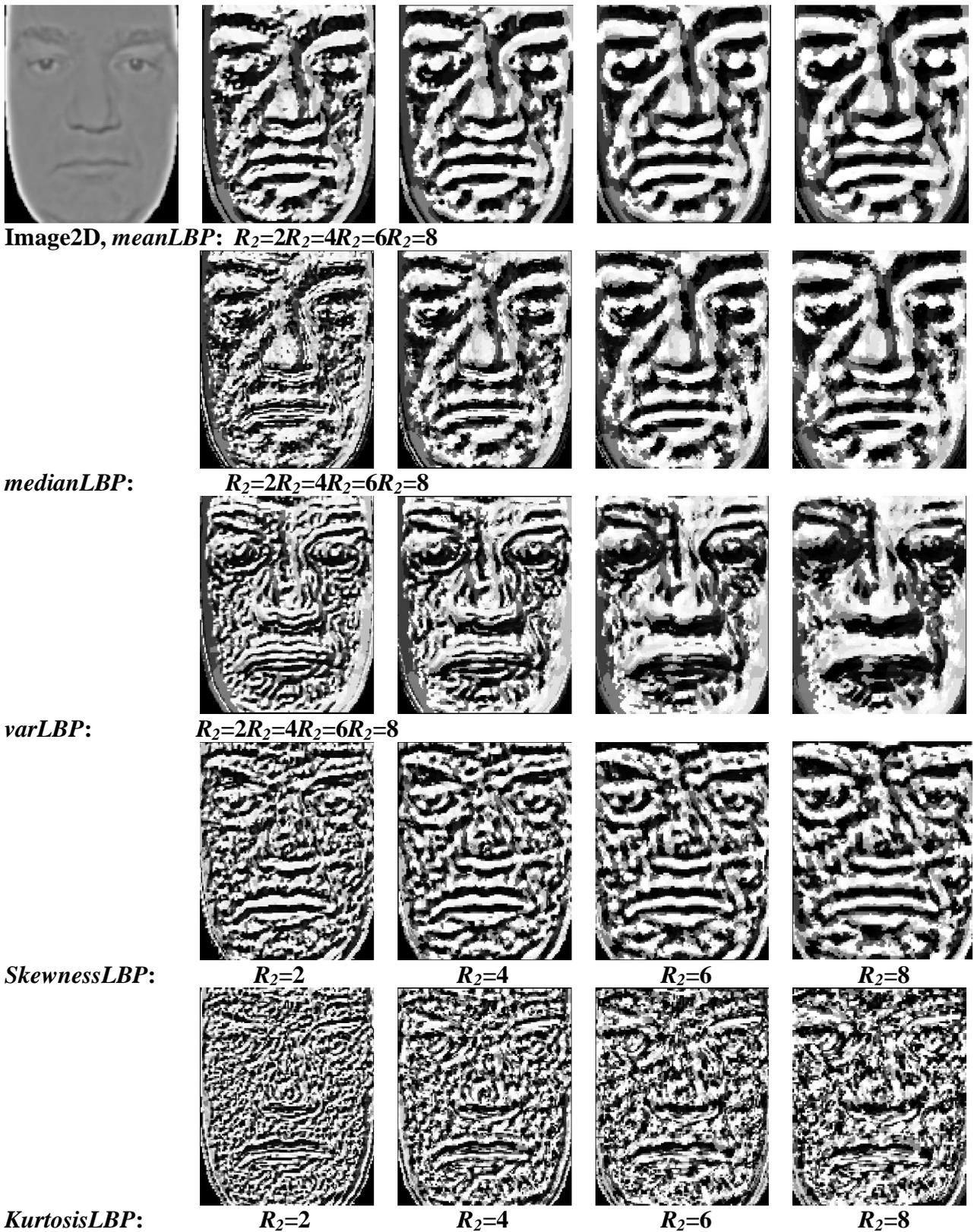


Figure 5.16: *Statistical LBP* pour les images de visage 2D.

5.3.3.1 Matrice exponentielle

La matrice exponentielle est largement utilisée dans des applications telles que la théorie de contrôle [74], la spectroscopie par résonance magnétique nucléaire [75], l'analyse de la chaîne de

Markov [359] et la réduction de la dimension [360]. Dans cette section, nous introduisons la définition et théorèmes de matrice exponentielle [361].

Définition 1:

Etant donné un ordre arbitraire de n matrices carrées A , son exponentielle est définie comme suit:

$$\exp(A) = I + A + \frac{A^2}{2!} + \dots + \frac{A^m}{m!} + \dots \quad (5.22)$$

Où I est une matrice unitaire de même taille que celle de A .

Théorème 1:

Compte tenu du fait que la matrice carrée A est finie, son exponentielle $\exp(A)$ est aussi une matrice finie.

Théorème 2:

Si la matrice carrée A commute avec B , c'est-à-dire $AB = BA$, alors:

$$\exp(A + B) = \exp(A) \exp(B) \quad (5.23)$$

Théorème 3:

Pour une matrice carrée inversible A , il existe l'inverse de $\exp(A)$ puis:

$$(\exp(A))^{-1} = \exp(-A) \quad (5.24)$$

Théorème 4:

Si T est une matrice inversible, alors:

$$\exp(T^{-1}AT) = T^{-1} \exp(A) T \quad (5.25)$$

Théorème 5:

Si $V = [v_1, v_2, \dots, v_n]$ sont les vecteurs propres de la matrice carrée A qui correspondent aux valeurs propres $[\lambda_1, \lambda_2, \dots, \lambda_n]$, alors $V = [v_1, v_2, \dots, v_n]$ sont également les vecteurs propres de la matrice $\exp(A)$ qui correspondent aux valeurs propres $[e^{\lambda_1}, e^{\lambda_2}, \dots, e^{\lambda_n}]$.

Ensuite, nous allons présenter la méthode EDA.

5.3.3.2 EDA

Pour le critère LDA, nous avons:

$$J(W) = \arg \max_W \frac{|W^T S_b W|}{|W^T S_w W|} = \arg \max_W \frac{|W^T (\Phi_b^T \Lambda_b \Phi_b) W|}{|W^T (\Phi_w^T \Lambda_w \Phi_w) W|} \quad (5.26)$$

Où $\Phi_b = (\Phi_{b_1}, \Phi_{b_2}, \dots, \Phi_{b_n})$ est la matrice des vecteurs propres de S_b qui correspond à valeurs propres $\Lambda_b = \text{diag}(\lambda_{b_1}, \lambda_{b_2}, \dots, \lambda_{b_n})$.

$\Phi_w = \text{diag}(\Phi_{w_1}, \Phi_{w_2}, \dots, \Phi_{w_n})$ est la matrice des vecteurs propres de S_w qui correspond aux valeurs propres $\Lambda_w = \text{diag}(\lambda_{w_1}, \lambda_{w_2}, \dots, \lambda_{w_n})$.

En général, la matrice S_w n'est pas une matrice "pleine", en particulier dans le cas d'un échantillon de petite taille. Cependant, l'information discriminante, qui correspond aux valeurs propres de S_w qui sont égales à 0, a le pouvoir le plus discriminant. Malheureusement, la LDA ne peut pas extraire des

informations discriminantes qui étaient contenues dans l'espace nul de S_w . Pour extraire ce type d'information discriminante, nous remplaçons λ_{w_i} , c'est à dire les valeurs propres de S_w par $\exp(\lambda_{w_i})$ et λ_{b_i} , c'est à dire les valeurs propres de S_b par $\exp(\lambda_{b_i})$. Notons que $\exp(\Lambda_b) = \text{diag}(e^{\lambda_{b_1}}, e^{\lambda_{b_2}}, \dots, e^{\lambda_{b_n}})$ et $\exp(\Lambda_w) = \text{diag}(e^{\lambda_{w_1}}, e^{\lambda_{w_2}}, \dots, e^{\lambda_{w_n}})$. Le critère de LDA est ensuite transformé en:

$$J(W) = \arg \max_W \frac{|W^T (\Phi_b^T \exp(\Lambda_b) \Phi_b) W|}{|W^T (\Phi_w^T \exp(\Lambda_w) \Phi_w) W|} = \arg \max_W \frac{|W^T \exp(S_b) W|}{|W^T \exp(S_w) W|} \quad (5.27)$$

D'après le théorème 3, la matrice $\exp(S_w)$ est une matrice pleine. Par conséquent, l'information discriminante qui était contenue dans l'espace nul de S_w peut-être extraite par le modèle (5.27), même si le problème de la taille petite de l'échantillon est impliqué.

Comme on le sait, l'orthogonalité est très importante pour l'analyse discriminante. L'avantage d'utiliser une transformation orthogonale est que les corrélations entre les caractéristiques de candidats sont décomposées de sorte que la signification de caractéristiques individuelles peut être évaluée de façon indépendante [362]. Par conséquent, on définit le critère EDA par application de la matrice de projection W (5.27) pour être orthogonale comme suit:

$$J(W) = \arg \max_{W^T W = I} \frac{|W^T \exp(S_b) W|}{|W^T \exp(S_w) W|} \quad (5.28)$$

Les avantages de l'EDA sont:

- 1) La matrice $\exp(S_w)$ est une matrice pleine. Par conséquent, l'information discriminante qui était contenue dans l'espace nul de S_w peut être extraite par l'expression (5.27).
- 2) La méthode du noyau est utilisée pour transformer les problèmes non linéaires d'origine en problèmes linéaires dans l'espace de caractéristiques transformé. Comme la méthode du noyau, la fonction exponentielle transforme les matrices de dispersion dans un autre espace.
- 3) L'objectif de la LDA est de maximiser la distance entre la classe et minimiser la distance à l'intérieur de cette classe. Ces distances peuvent être calculées par la trace de matrices de diffusion correspondantes:

$$\text{trace}(S_b) = \lambda_{b_1} + \lambda_{b_2} + \dots + \lambda_{b_n}$$

$$\text{et } \text{trace}(S_w) = \lambda_{w_1} + \lambda_{w_2} + \dots + \lambda_{w_n}$$

La plus grande valeur du rapport $\frac{\lambda_{b_i}}{\lambda_{w_i}}$ a le plus fort pouvoir discriminant. Selon le théorème 5, nous avons:

$$\text{trace}(\exp(S_b)) = \exp(\lambda_{b_1}) + \exp(\lambda_{b_2}) + \dots + \exp(\lambda_{b_n}).$$

$$\text{trace}(\exp(S_w)) = \exp(\lambda_{w_1}) + \exp(\lambda_{w_2}) + \dots + \exp(\lambda_{w_n}).$$

Et:

$$\frac{\exp(\lambda_{b_i})}{\exp(\lambda_{w_i})} > \frac{\lambda_{b_i}}{\lambda_{w_i}} \quad (5.29)$$

Comme on le voit, il existe une différence d'échelle de diffusion entre les distances intra-classes et inter-classes. En effet, la diffusion de l'échelle de la distance entre la classe est plus grande que la distance à l'intérieur de la classe. Par conséquent, la marge entre les différentes classes est agrandie.

On peut définir donc l'algorithme EDA comme suit:

Entrée: Ensemble de données X .

Sortie: matrice de transformation W

1. Calcul des matrices S_b , S_w , $\exp(S_b)$ et $\exp(S_w)$.
2. Calcul des vecteurs propres $\{v_i\}$ et les valeurs propres correspondantes $\{\lambda_i\}$ de $(\exp(S_w))^T \exp(S_b)$.
3. Tri des vecteurs propres $W = \{v_i\}$ selon l'ordre décroissant des valeurs propres $\{\lambda_i\}$.
4. Rendre orthogonale la matrice de transformation W .

5.3.4 WCCN (Within Class Covariance Normalization) et mesure de similarité

WCCN a été utilisée principalement en reconnaissance vocale et a d'abord été introduite dans [352]. Il se trouve dans [363] que la meilleure approche consiste à projeter i-vecteurs réduits par la LDA à un nouveau sous-espace indiqué par la racine carrée de l'inverse de la matrice de covariance intra-classe suivante:

$$K = \sum_{i=1}^L \frac{1}{n_i} \sum_{j=1}^{n_i} (W^T x_j^i - \widetilde{m}^i)(W^T x_j^i - \widetilde{m}^i)^T \quad (5.30)$$

Où: $\widetilde{m}^i = \frac{1}{n_i} \sum_{j=1}^{n_i} W x_j^i$ et W est la matrice de projection EDA trouvée par l'expression 5.28. La matrice B de projection de la méthode WCCN peut être obtenue par la décomposition de *Cholesky* de l'inverse de W ($W^{-1} = BB^T$). WCCN pour déclasser l'effet de la direction de haute inter variabilité et améliorer la discrimination de la similarité cosinus [381].

Nous utilisons la mesure de similarité cosinus après analyse discriminante du fait de sa relation avec la règle de décision de *Bayes*. Le classifieur de *Bayes* est l'optimal pour minimiser l'erreur de classification [352]. La méthode permet de calculer le score cosinus entre les deux régions caractéristiques locales (x_1 et x_2) dans l'espace de projection EDA + WCCN:

$$S_{cos}(x_1, x_2) = \frac{(B^T W^T x_1)^T (B^T W^T x_2)}{\|B^T W^T x_1\| \|B^T W^T x_2\|} \quad (5.31)$$

5.4 Résultats

L'efficacité des méthodes et l'approche proposée sont évaluées dans cette section. Tout d'abord, nous décrivons brièvement les ensembles de données sur lequel les expériences ont été réalisées. Nous évaluons ensuite la vérification du visage uni-modal 2D et 3D pour différentes méthodes d'extraction de caractéristiques binaires locales à différentes échelles. Enfin, la vérification du visage multimodale et multi-descripteurs est effectuée et les résultats sont comparés avec ceux de la littérature scientifique.

5.4.1 Bases de données

Nous validons notre méthode proposée en la comparant avec celles utilisant sur les bases de données FRGC v2 et CASIA 3D. La base de données CASIA 3D est présentée dans la section 4.2.4.1.

5.4.1.1 La base de données FRGC v2

La base de données FRGC v2 (Face Recognition Grand Challenge version 2) comprend 4007 images du visage de 466 personnes distinctes, avec jusqu'à 22 images par sujet. Les images ont été acquises avec «Minolta Vivid 910» [55]. Les deux informations de couleur et coordonnées 3D ne sont pas parfaitement capturées simultanément car la bande de laser nécessite quelques secondes pour traverser le visage. La résolution de l'appareil photo Minolta est de taille 640×480, ce qui donne

environ 300000 points d'échantillonnage possibles. Le nombre de points 3D sur une image frontale du visage prises par la caméra Minolta est d'environ 112000. Des points supplémentaires proviennent de cheveux, vêtements et fond des objets. Des exemples d'images de ce capteur peuvent être vus sur la figure 5.17.

Cependant, la résolution des visages varie parce qu'ils ont été scannés à différentes distances du scanner en utilisant peut-être d'autres lentilles. Plus de 40 % des visages ont des occlusions capillaires. Il y a aussi de fortes variations d'expression (figure 5.18) et différentes conditions d'éclairage obtenues dans des environnements non contrôlés. Certains artefacts sont présents lors de l'acquisition tels que des images déformées, l'absence d'un nez et les trous autour des yeux et de la bouche (figure 5.18). Cette base de données contient des variations complexes qui sont difficiles pour n'importe quel algorithme.

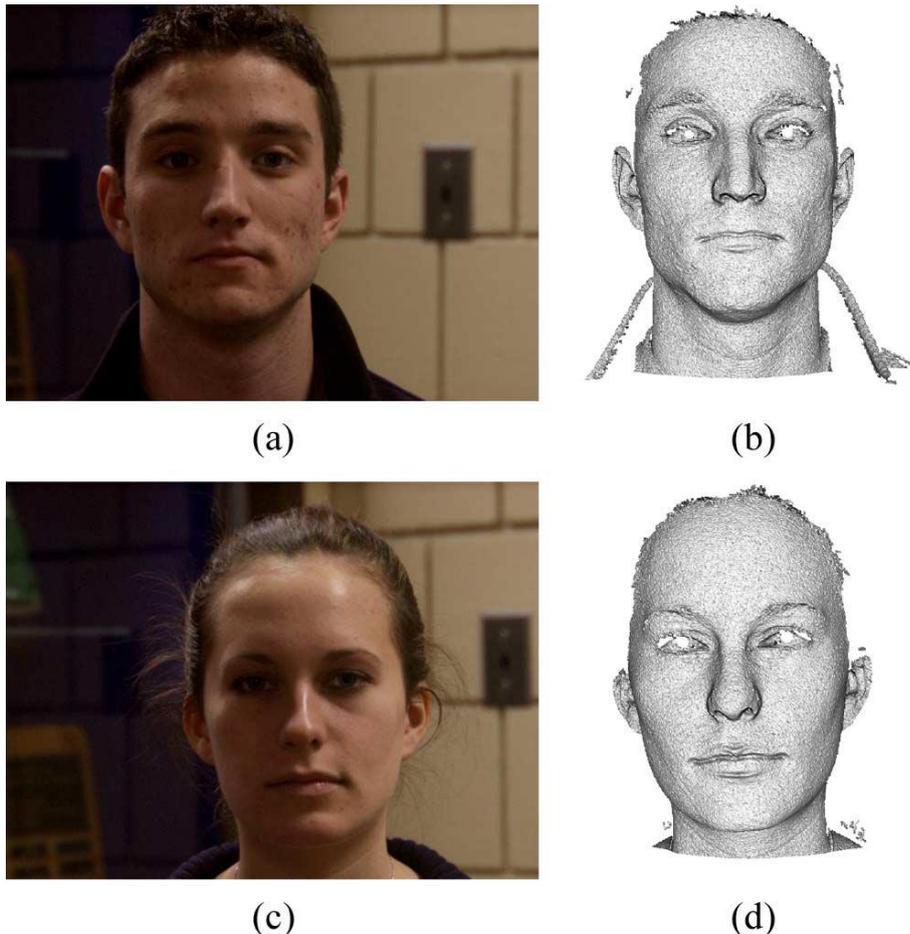
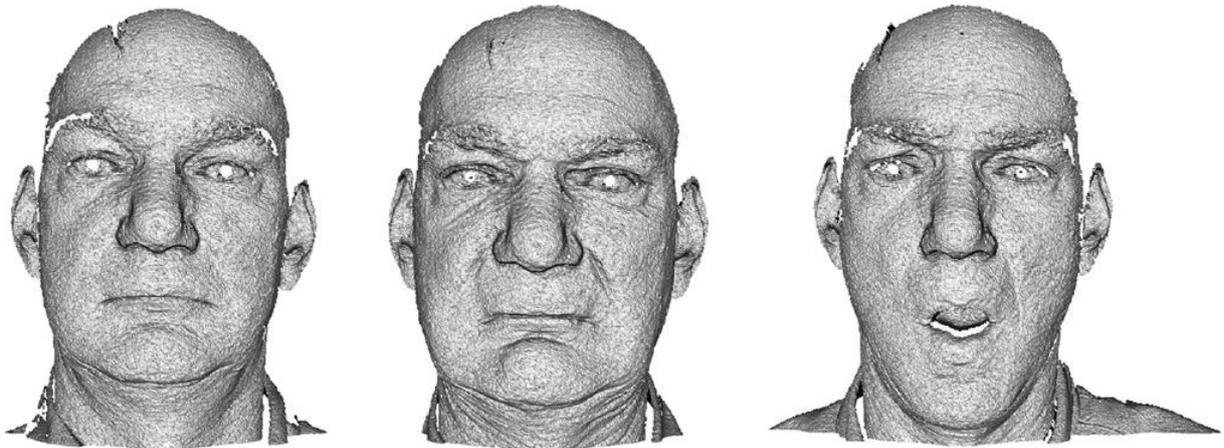


Figure 5.17: Echantillons d'images capturées avec Minolta Vivid 910, (a) et (c) images 2D, (b) et (d) images 3D.



(a)

(b)

(c)

Figure 5.18: Exemples d'images. (a) Image neutre. (b) Petite expression. (c) Grande expression.



(a)

(b)



(c)

(d)

Figure 5.19: Exemples d'images contenant des artefacts dans l'ensemble FRGC v2.

Les données pour l'expérience 3 de la base de données FRGC v2 ont été divisées en deux parties: formation et validation. L'ensemble de la formation contient 943 images de 266 sujets qui sont utilisées pour la formation d'EDA et WCCN. L'ensemble de validation contient 4007 images de 466 sujets. La base FRGC v2 donne également un ensemble de 4007 images de formation et un ensemble de 4007 images d'évaluation. Selon le protocole de FRGC, chaque visage de formation doit être comparé à chaque visage de validation, ce qui donne 16 millions de scores de similarité. La courbe ROC trace le taux de vérification (Verification Rate) par rapport au taux de fausse acceptation (False Acceptance Rate) qui est généré en utilisant l'outil d'évaluation BEE «Biometric Experimentation

Environment». L'outil BEE génère trois courbes ROC (ROC I, II et III) correspondant aux images acquises dans les semestres, dans une année et entre les semestres respectivement. Nous utilisons le protocole "neutre contre tout" pour évaluer la performance en cas de variation d'expression car il est très utilisé dans la littérature.

5.4.2 Vérification uni-modale de visage 2D et 3D

Dans cette section, nous comparons la vérification uni-modale de visage 2D et 3D dans les bases de données FRGC v2 et CASIA 3D et testons quatre méthodes d'extraction de caractéristiques binaires locales dans les données 2D et 3D.

Le tableau 5.4 montre le taux de vérification (TV) à 0,001 TFA sur la base de trois masques standards de la FRGC v2, appelés ROC I, ROC II et ROC III pour la vérification uni-modale du visage 2D et 3D. Le tableau 5.5 montre le taux d'égale d'erreur (TEE) dans l'ensemble d'évaluation et le taux de vérification (VR = 100 - TFA - TFR) dans l'ensemble de test pour la base de données CASIA 3D. TFR est le taux de faux rejet et TFA le taux de fausse acceptation.

- Pour la LBP nous utilisons : ($P = 10, R = 2, 4, 6$ et 8) pour le visage 3D et la LBP^{u2} avec ($P = 8, R = 2, 4, 6$ et 8) pour le visage 2D.
- Pour la PLQ nous utilisons les fenêtres de taille $M = 5, 7, 13$ et 17 .
- Pour BSIF originale (BSIFo) et BSIF nouvelle (BSIFn) nous utilisons des filtres de taille $l = 5, 9, 11, 13, 15$.
- Pour *Statistical LBP* nous utilisons $R_1 = 3, P_1 = 24, P_2 = 10$ et $R_2 = 2, 4, 6, 8$.
- Le nombre des vecteurs propres pour l'EDA est égal à 400 pour la base de données FRGC v2 et 100 pour CASIA 3D.

Les tableaux 5.4 et 5.5 montrent plusieurs observations et conclusions présentées comme ci-dessous.

- Pour les deux bases de données utilisées, nous avons obtenu à peu près les mêmes remarques et résultats.
- La stratégie multi-échelles est toujours meilleure qu'une stratégie à une seule échelle, pour tous les descripteurs binaires locaux utilisés et pour la vérification uni-modale de visage 2D et 3D.
- Tous les descripteurs: *mean LBP*, *var LBP*, *skewness LBP* et *kurtosis LBP* sont meilleurs que le descripteur LBP pour toute échelle.
- Le nouveau descripteur BSIF (BSIFn) est meilleur que celui de BSIF original (BSIFo) pour la vérification de visage 2D et 3D et pour toute échelle et multi-échelles. MSBSIFn est le meilleur descripteur de vérification de visage 3D avec la génération des filtres par les images 3D.
- *Statistical LBP* proposée avec $R_2 = 4$ est meilleure que MSLBP et MSLPQ pour la vérification de visage 3D.
- Notre descripteur *statistical LBP* proposé donne la meilleure performance pour la vérification de visage 2D.
- Pour tous les descripteurs binaires locaux étudiés, nous remarquons clairement qu'il y a une petite différence de performance pour les trois ROC sur la reconnaissance faciale en 3D, mais il y a de grandes différences de performances pour la reconnaissance faciale 2D. Les données 3D sont donc plus stables que les données 2D pour la base de données FRGC v2.

Tableau 5.4: Taux de vérification uni-modale 2D et 3D de visage à 0,001 TFA pour le protocole standard ROC I, II et III (FRGC v2).

Extraction de caractéristiques locales binaires		Visage 2D			Visage 3D		
		ROC I	ROC II	ROC III	ROC I	ROC II	ROC III
Local	$LBP_{R=2}$	79.00	70.37	60.64	84.95	83.26	81.60
Binary	$LBP_{R=4}$	82.43	74.02	64.86	85.59	84.14	82.72
Patterns	$LBP_{R=6}$	73.02	65.49	57.13	84.96	83.46	381.96
(LBP)	$LBP_{R=8}$	69.16	61.78	53.25	83.47	81.84	79.90
	MLBP	85.76	74.40	64.69	86.33	84.94	83.35
Local	$LPQ_{M=5}$	89.41	78.77	68.37	89.96	88.91	87.79
Phase	$LPQ_{M=7}$	90.59	80.61	70.56	91.88	90.89	89.83
Quantization	$LPQ_{M=13}$	88.43	78.22	67.75	91.52	90.47	89.51
(LPQ)	$LPQ_{M=17}$	85.26	74.32	63.57	90.50	89.45	88.09
	MLPQ	90.92	80.83	70.75	92.28	91.25	90.29
Binarized	$BSIF_{O=7}$	86.31	75.66	63.67	89.35	88.41	87.27
Statistical	$BSIF_{O=9}$	86.49	75.82	63.67	90.62	89.72	88.51
Image	$BSIF_{O=11}$	86.49	75.21	63.70	91.53	90.63	89.66
Features	$BSIF_{O=13}$	85.69	74.33	62.04	91.28	90.46	89.40
original	$BSIF_{O=15}$	84.46	72.83	60.45	90.77	89.65	88.33
(BSIFo)	MBSIFo	86.60	76.85	64.51	91.66	90.68	89.75
Binarized	$BSIFn_{=7}$	86.51	75.62	64.84	90.00	89.31	88.69
Statistical	$BSIFn_{=9}$	87.21	75.89	64.17	92.16	91.50	90.70
Image	$BSIFn_{=11}$	86.38	74.98	63.26	92.83	92.15	91.21
Features	$BSIFn_{=13}$	86.02	74.92	63.53	93.16	92.57	91.80
nouveau	$BSIFn_{=15}$	85.21	73.45	61.31	92.97	92.43	91.92
(BSIFn)	MSBSIFn	87.59	76.64	65.20	93.20	92.66	91.96
	<i>mean LBP</i> _{R2=2}	86.34	79.77	73.64	87.96	86.49	84.75
	<i>mean LBP</i> _{R2=4}	85.29	78.65	71.91	92.45	91.27	90.03
	<i>mean LBP</i> _{R2=6}	81.63	75.79	69.81	89.50	88.25	87.06
	<i>mean LBP</i> _{R2=8}	79.78	73.58	67.33	90.97	89.96	88.68
	<i>MSmean LBP</i>	88.49	80.33	71.92	92.42	91.49	90.35
	<i>var LBP</i> _{R2=2}	86.30	79.74	73.65	91.11	90.05	88.91
	<i>var LBP</i> _{R2=4}	85.30	78.67	71.92	91.58	90.40	89.10
	<i>var LBP</i> _{R2=6}	81.72	75.74	69.80	91.60	90.79	89.84
	<i>var LBP</i> _{R2=8}	79.80	73.72	67.31	90.91	89.91	88.72
	<i>MSvar LBP</i>	88.49	80.34	72.08	92.49	91.61	90.41
Statistical	<i>Skewness LBP</i> _{R2=2}	84.51	75.70	66.63	89.22	87.95	86.57
Local	<i>Skewness LBP</i> _{R2=4}	85.33	77.15	68.73	90.92	89.77	88.59
Binary	<i>Skewness LBP</i> _{R2=6}	82.97	74.68	66.337	90.80	89.85	88.78
Patterns	<i>Skewness LBP</i> _{R2=8}	79.23	70.43	61.227	91.36	90.52	89.52
	<i>MSSkewness LBP</i>	87.81	78.67	69.17	91.60	90.63	89.50
	<i>Kurtosis LBP</i> _{R2=2}	75.50	66.91	57.52	86.46	84.93	83.18
	<i>Kurtosis LBP</i> _{R2=4}	76.11	67.75	58.49	88.79	87.54	86.09
	<i>Kurtosis LBP</i> _{R2=6}	75.50	67.32	58.39	88.45	87.55	86.46
	<i>Kurtosis LBP</i> _{R2=8}	73.60	64.73	55.34	87.68	86.58	85.34
	<i>MSkurtosis LBP</i>	84.94	76.86	67.99	89.80	88.64	87.46
	<i>meadian LBP</i> _{R2=2}	83.05	75.55	67.96	84.16	82.41	80.44
	<i>meadian LBP</i> _{R2=4}	83.34	75.33	67.40	84.16	82.47	80.75
	<i>meadian LBP</i> _{R2=6}	81.75	74.34	67.20	83.81	82.13	80.19
	<i>meadian LBP</i> _{R2=8}	79.02	71.97	65.51	83.24	81.44	79.44
	<i>MSmeadian LBP</i>	86.93	77.41	67.47	85.74	84.40	82.7
	<i>Statistical LBP</i> _{R2=2}	91.35	90.10	88.77	91.18	82.70	74.19
	<i>Statistical LBP</i> _{R2=4}	92.44	91.49	90.44	90.79	82.83	74.97
	<i>Statistical LBP</i> _{R2=6}	91.84	91.00	90.10	89.70	81.94	74.24
	<i>Statistical LBP</i> _{R2=8}	91.84	91.00	90.10	88.31	80.63	72.81
	<i>MSStatistical LBP</i>	92.18	91.18	90.23	90.82	83.14	75.66

Tableau 5.5: Vérification uni-modale 3D et 2D de visage (CASIA 3D).

Extraction de caractéristiques locales binaires		Visage 2D		Visage 3D	
		TEE éval	TV test	TEE éval	TV test
Local	$LBP_{R=2}$	1.02	91.84	2.40	91.16
Binary	$LBP_{R=4}$	0.95	91.57	2.95	92.95
Patterns	$LBP_{R=6}$	1.95	92.45	3.80	93.26
(LBP)	$LBP_{R=8}$	1.04	92.14	3.75	93.36
	MLBP	0.80	93.60	2.61	94.10
Local	$LPQ_{M=5}$	0.78	94.42	1.04	95.20
Phase	$LPQ_{M=7}$	1.03	93.20	1.16	95.00
Quantization	$LPQ_{M=13}$	0.80	93.72	0.78	96.00
(LPQ)	$LPQ_{M=17}$	0.75	95.54	0.98	96.05
	MSLPQ	0.39	96.06	0.62	96.97
Binarized	$BSIF_{O_{f=7}}$	2.85	92.03	2.21	94.70
Statistical	$BSIF_{O_{f=9}}$	2.47	92.51	2.01	95.10
Image	$BSIF_{O_{f=11}}$	2.98	92.98	2.36	95.48
Features	$BSIF_{O_{f=13}}$	2.35	92.78	2.44	95.59
original	$BSIF_{O_{f=15}}$	2.37	91.86	1.63	95.38
(BSIFo)	MSBSIFo	1.91	94.83	1.52	95.87
Binarized	$BSIFn_{f=7}$	0.99	93.92	0.60	96.59
Statistical	$BSIFn_{f=9}$	1.19	94.96	0.56	96.62
Image	$BSIFn_{f=11}$	1.61	94.32	0.59	96.95
Features	$BSIFn_{f=13}$	1.18	95.55	0.75	96.96
nouveau	$BSIFn_{f=15}$	1.18	94.36	0.63	96.75
(BSIFn)	MSBSIFn	0.64	96.10	0.55	97.27
Statistical	$mean LBP_{R2=4}$	0.37	94.57	0.84	97.14
Local	$var LBP_{R2=4}$	0.36	94.46	0.95	97.14
Binary	$skewness LBP_{R2=4}$	0.24	95.86	1.39	96.49
Patterns	$kurtosis LBP_{R2=4}$	0.77	96.48	0.99	96.32
	$meadian LBP_{R2=4}$	0.41	94.96	1.16	96.48
	$Statistical LBP_{R2=4}$	0.19	96.75	0.81	97.24

Les figures 5.20, 5.21, 5.22, 5.23, 5.24, 5.25, 5.26 et 5.27 présentent les trois courbes ROC de vérification uni-modale de visage 3D et 2D (la base de données FRGC v2)

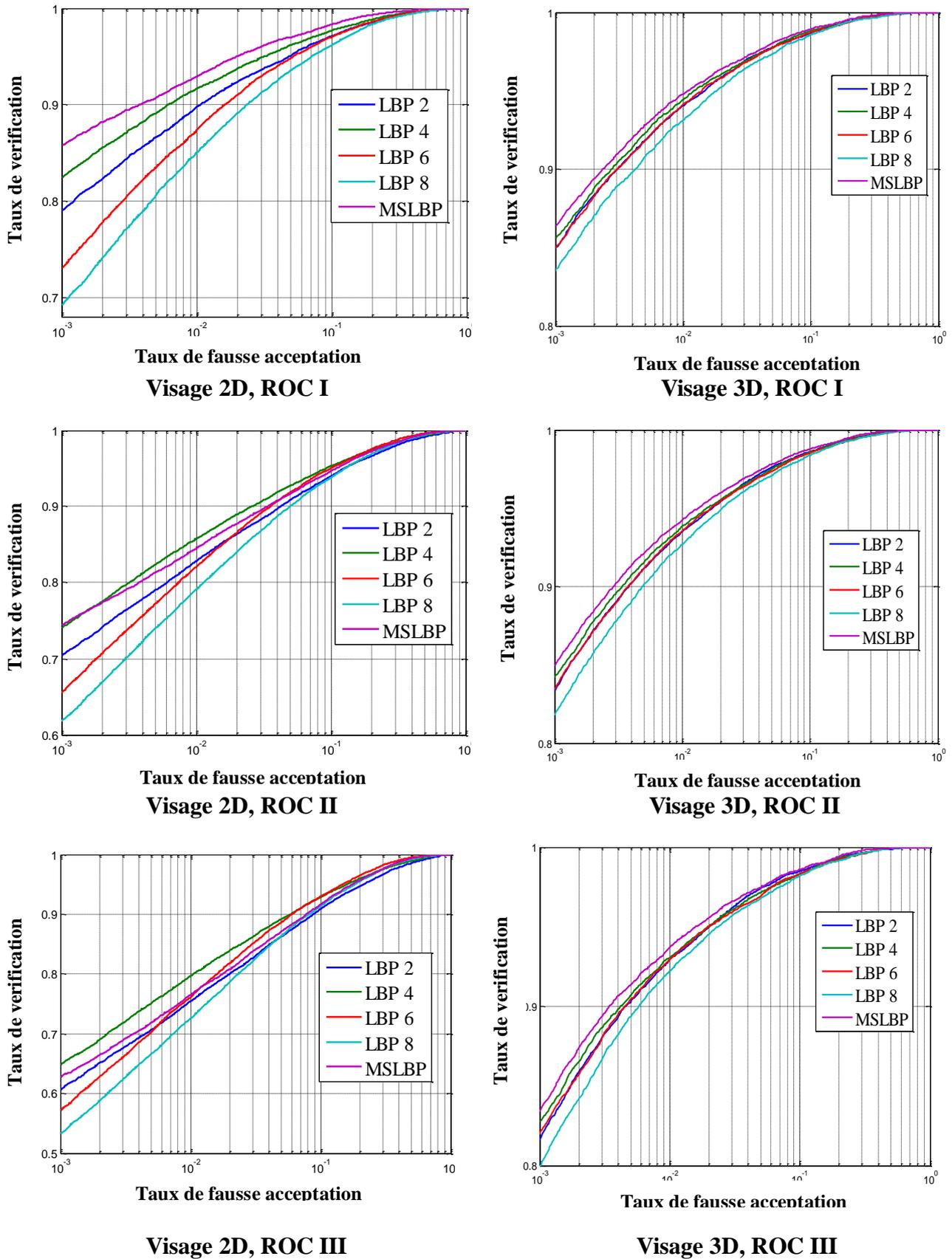
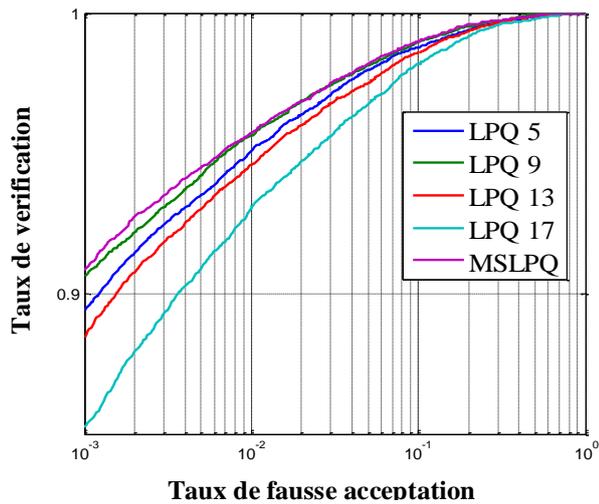
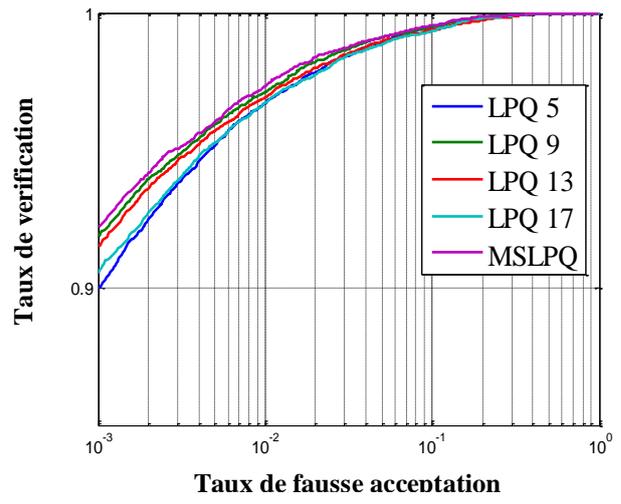


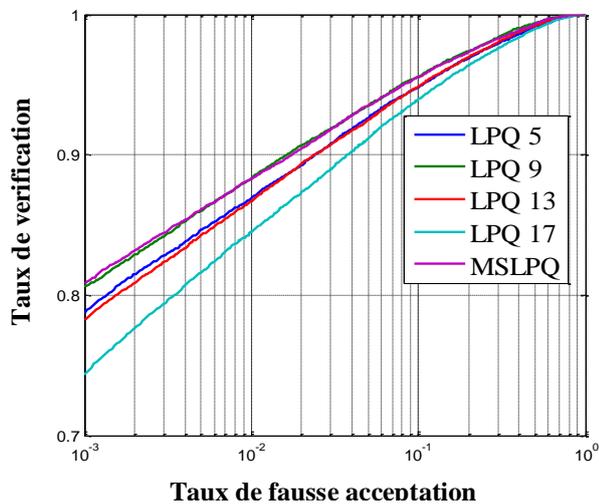
Figure 5.20: Courbes ROC pour le descripteur LBP.



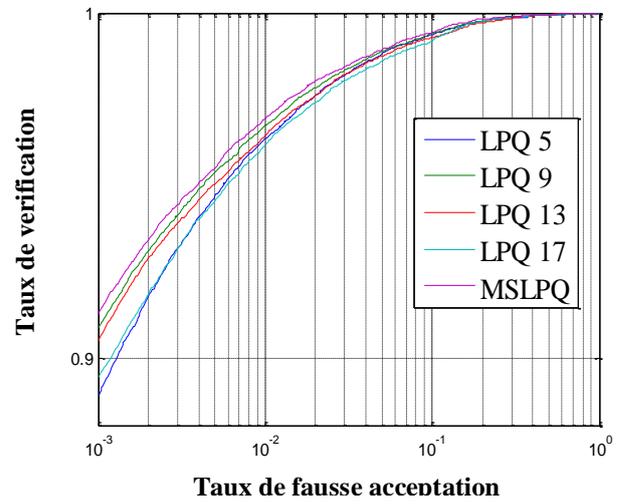
Visage 2D, ROC I



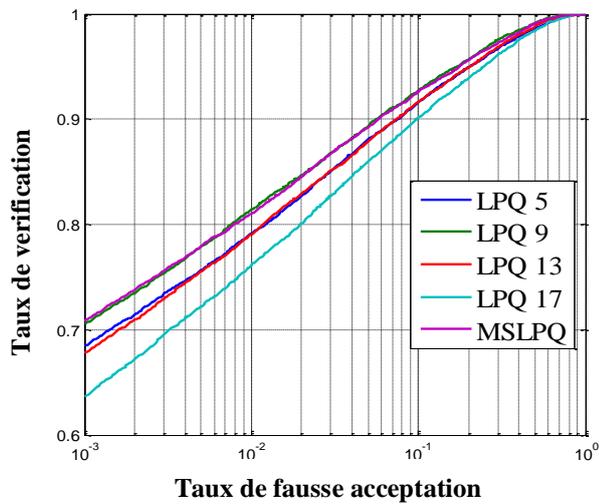
Visage 3D, ROC I



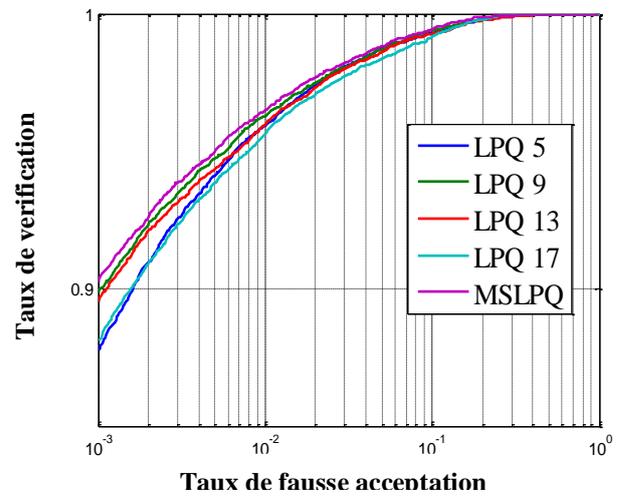
Visage 2D, ROC II



Visage 3D, ROC II



Visage 2D, ROC III



Visage 3D, ROC III

Figure 5.21: Courbes ROC pour le descripteur LPQ.

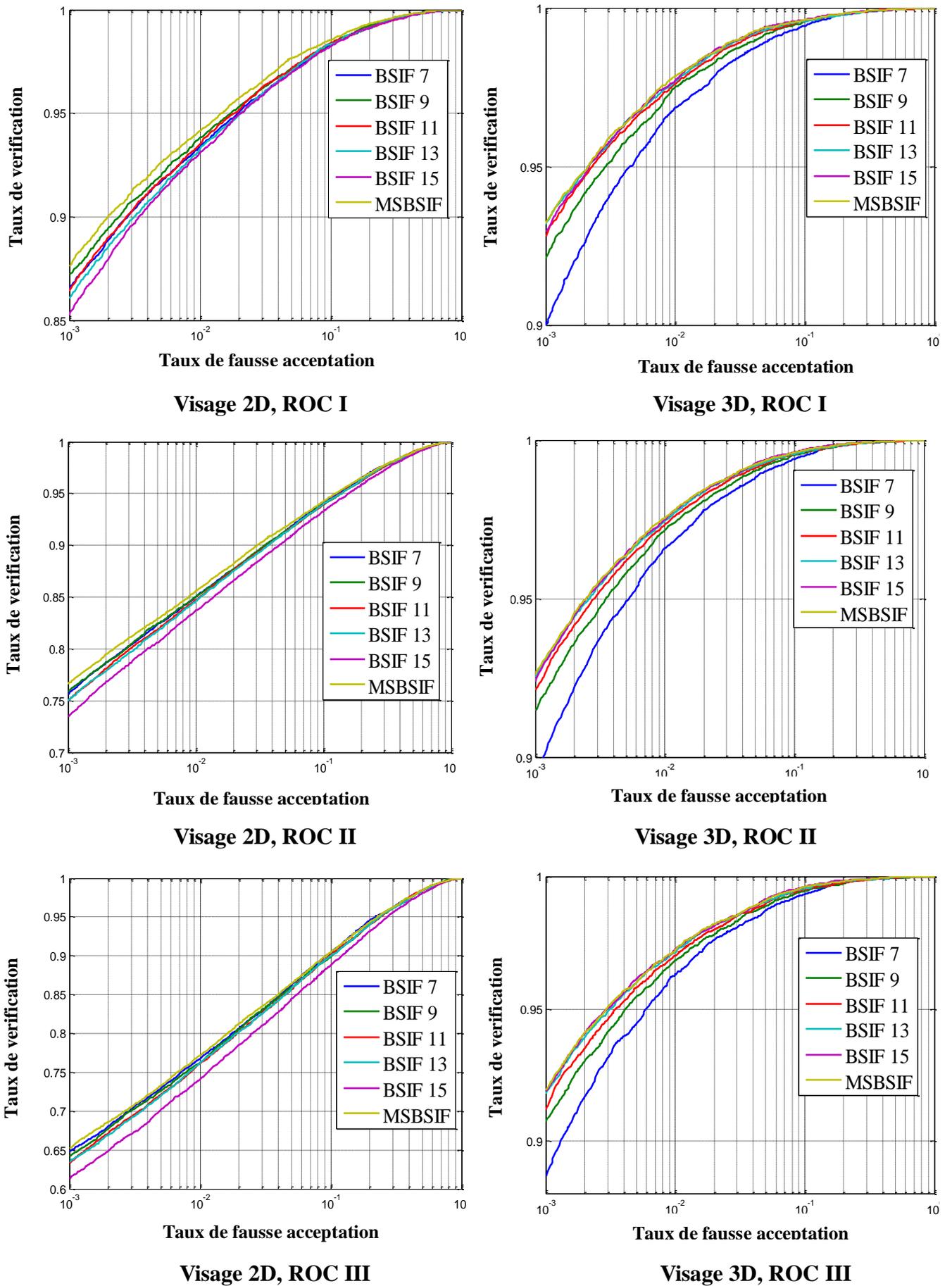


Figure 5.22: Courbes ROC pour le descripteur BSIFo.

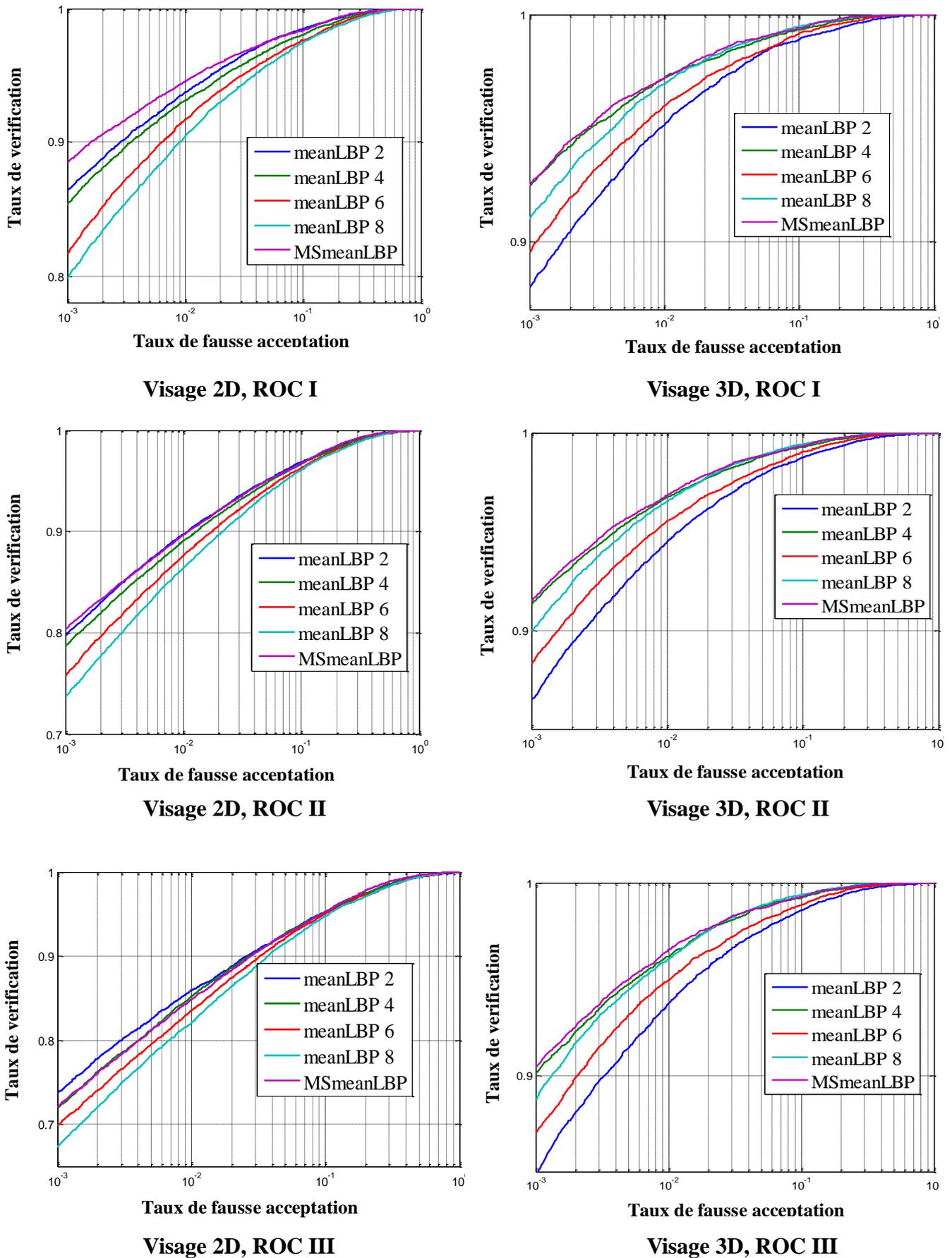


Figure 5.23: Courbes ROC pour le descripteur *mean LBP*.

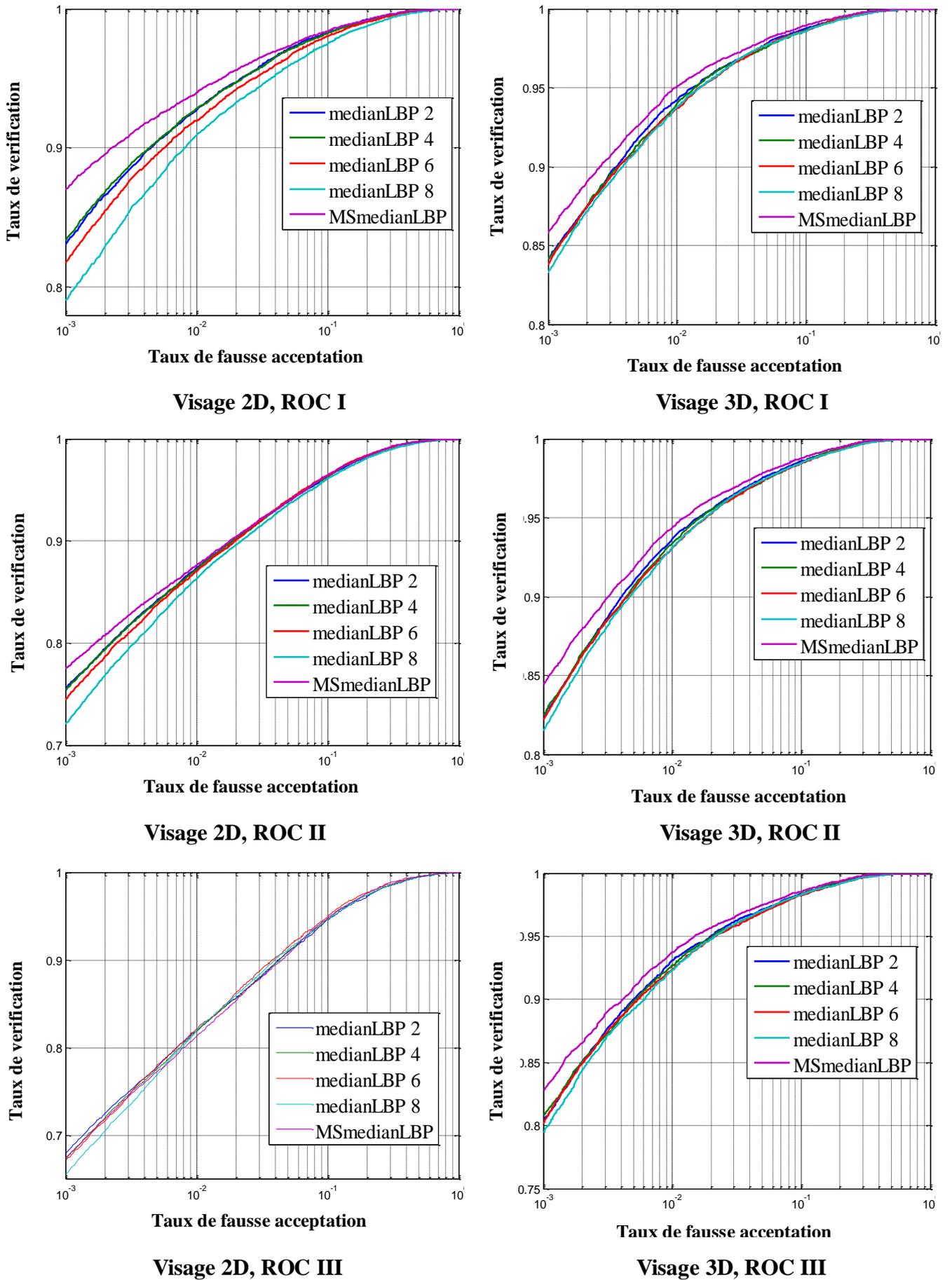


Figure 5.24: Courbes ROC pour le descripteur *median LBP*.

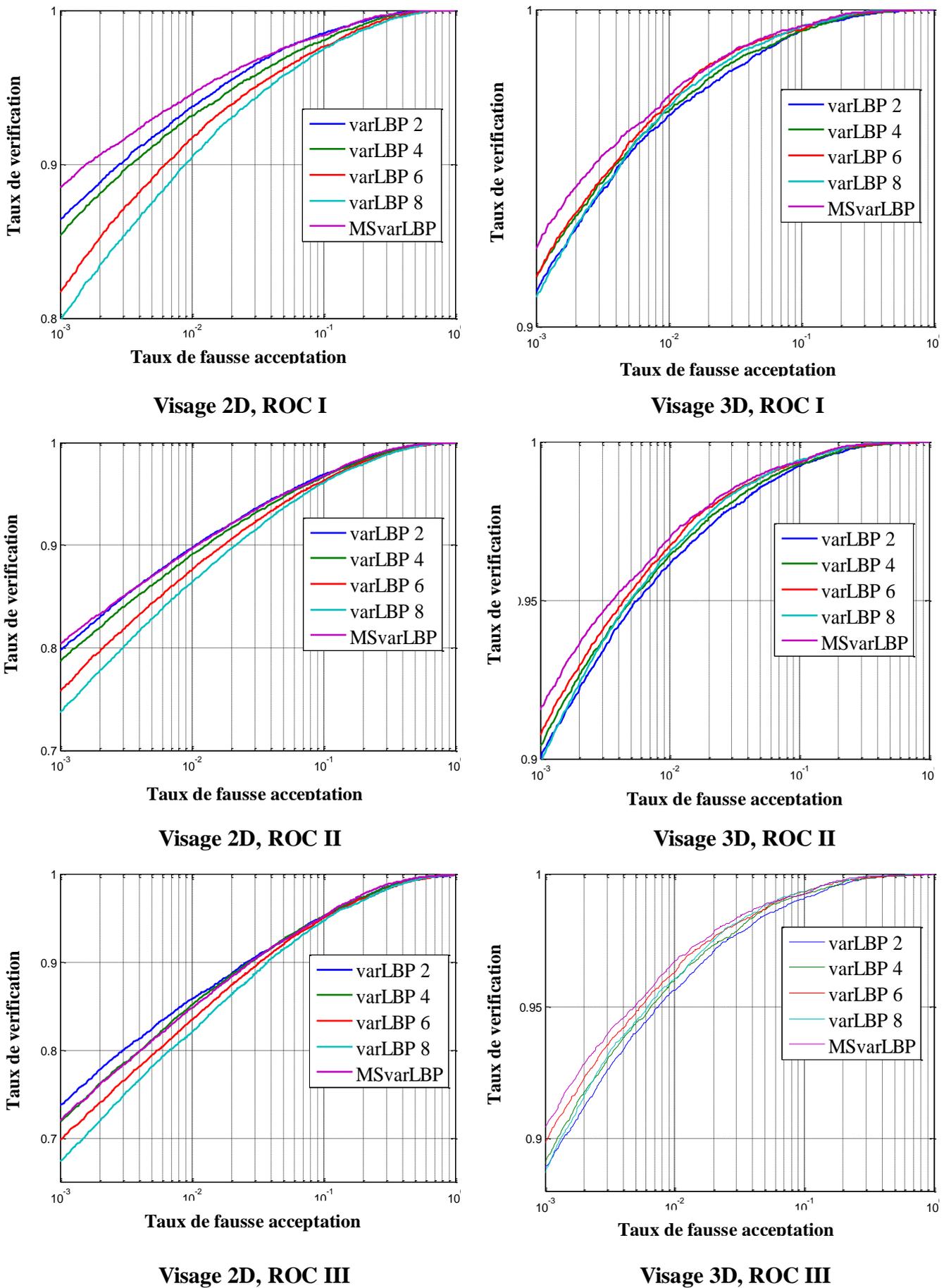


Figure 5.25: Courbes ROC pour le descripteur *var LBP*.

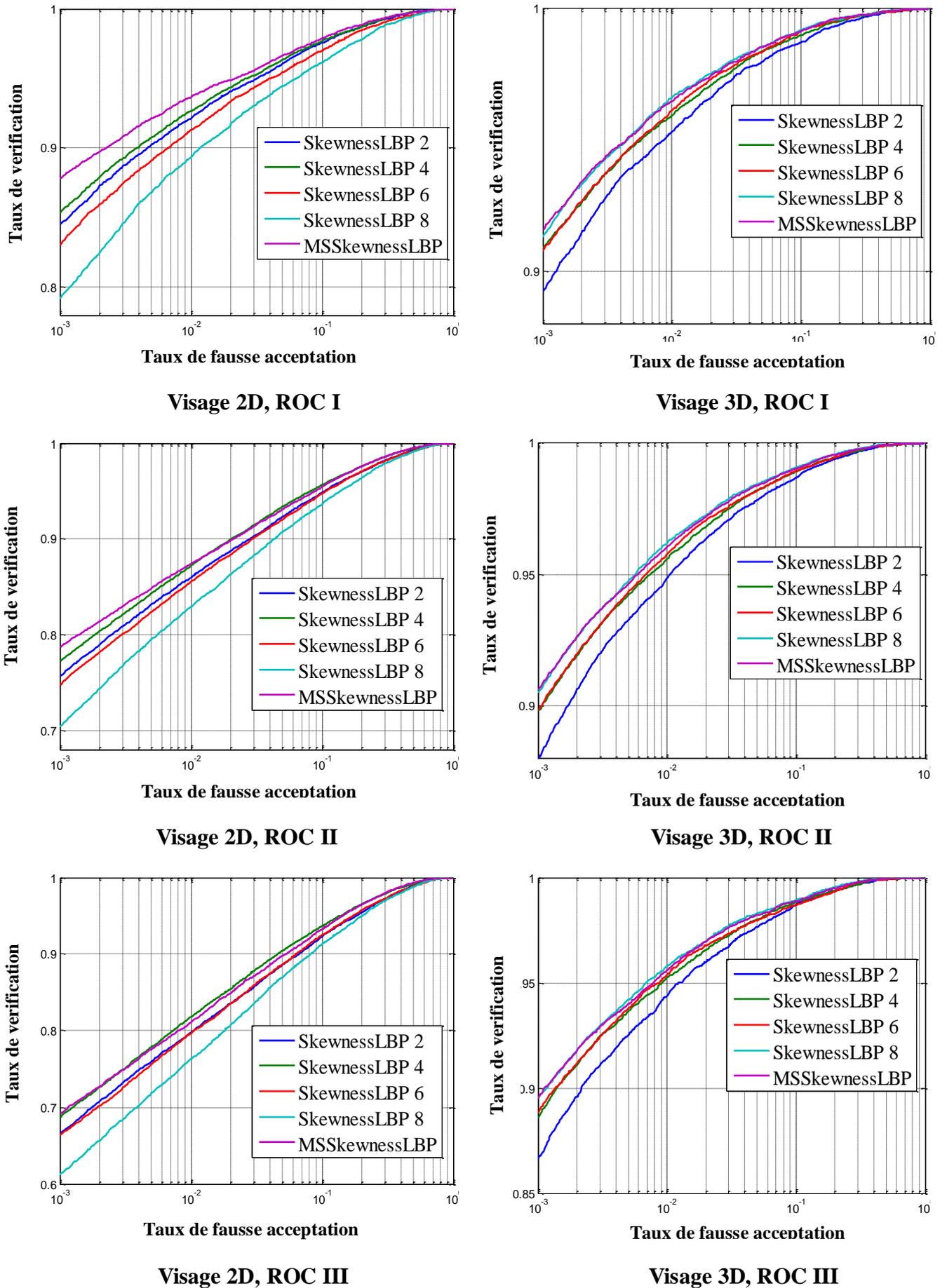


Figure 5.26: Courbes ROC pour le descripteur *Skewness LBP*.

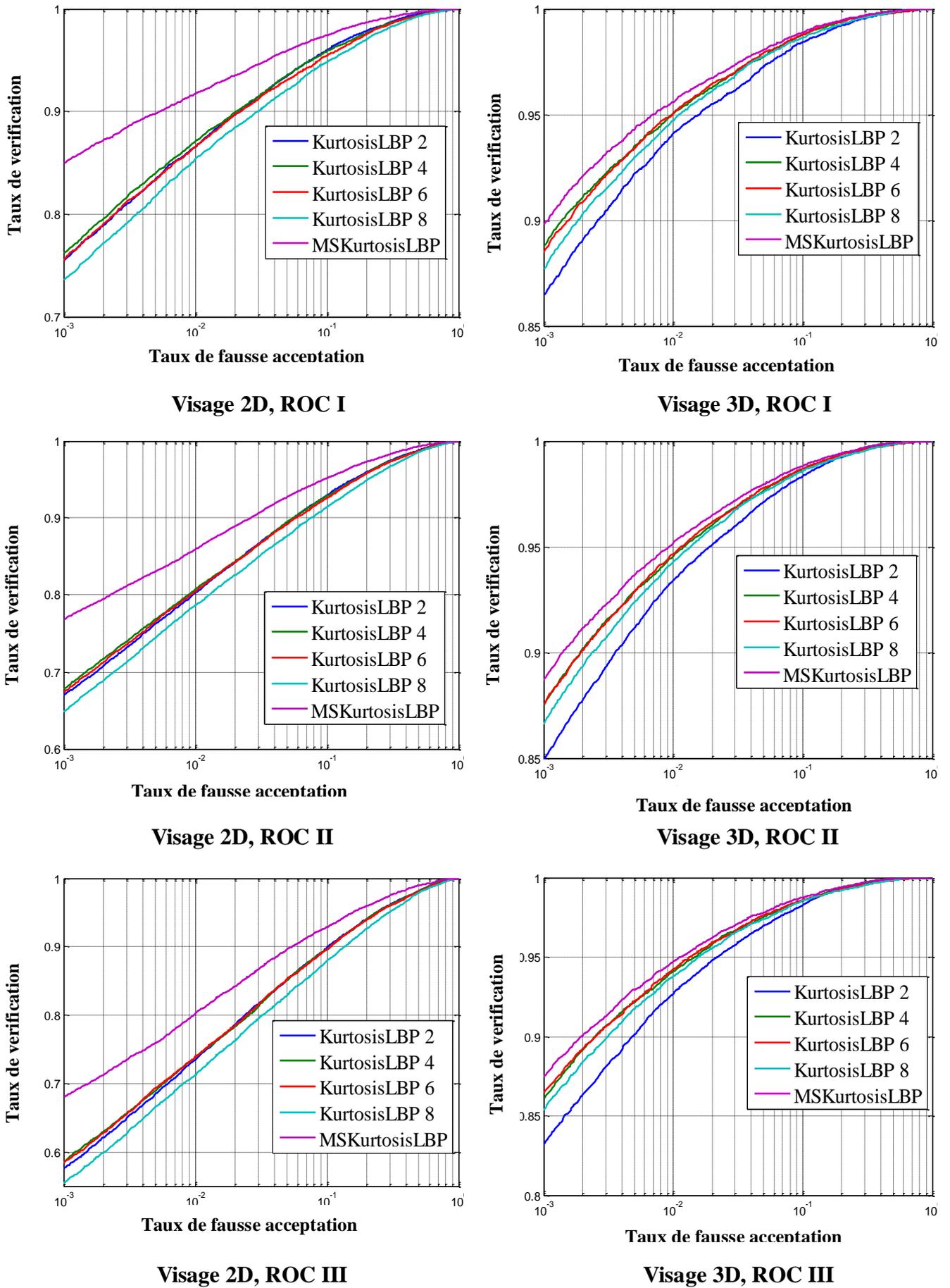


Figure 5.27: Courbes ROC pour le descripteur *Kurtosis LBP*.

5.4.3 Vérification de visage multi-descripteurs et multimodales (2D + 3D)

Dans cette section, nous avons présenté les descripteurs multi-échelles locaux binaires (MSLBP, MSLPQ et MSBSIF) et notre descripteur proposé *statistical LBP* ($R_1 = 3, P_1 = 24, R_2 = 4, P_2 = 10$) pour la vérification multimodale de visage (2D+3D). Nous présentons également la fusion des différents descripteurs et une comparaison de notre travail avec les travaux de la littérature scientifique en reconnaissance de visage 2D et 3D. Le tableau 5.6 montre les taux de vérification à 0,001 TFA pour les trois masques standards de la base de données FRGC v2 (ROC I, ROC II et ROC III). Nous considérons aussi le protocole "neutre contre tout" pour évaluer la performance en cas de variation de l'expression. Le tableau 5.7 montre le TEE dans l'ensemble d'évaluation et le taux de vérification (TV) dans l'ensemble de tests pour la base de données CASIA.

Les figures 5.28 et 5.29 présentent les courbes ROC III et les courbes ROC pour le protocole "neutre contre tout", respectivement pour la vérification multi-descripteurs et multi-modalités.

Tableau 5.6: Taux de vérification multimodale du visage à 0,001 TFA pour ROC I, ROC II, ROC III et ROC pour le protocole "neutre contre tout".

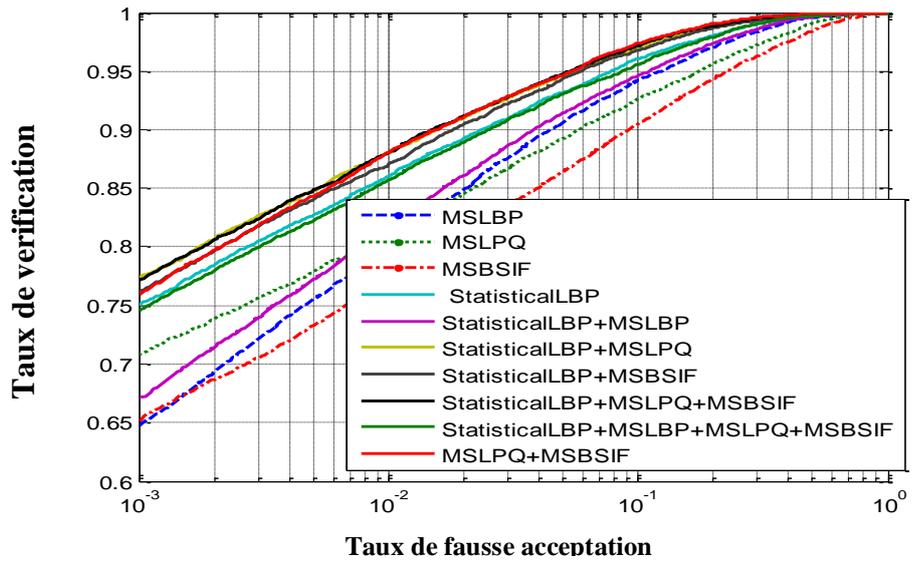
	Visage 2D				Visage 3D				Visage 2D+3D			
	ROC I	ROC II	ROC III	neutre contre toute	ROC I	ROC II	ROC III	neutre contre toute	ROC I	ROC II	ROC III	neutre contre toute
1- MSLBP	85.76	74.40	64.69	72.97	86.33	84.94	83.35	88.12	89.87	86.43	83.46	94.11
	90.92	80.83	70.75	87.77	92.28	91.25	90.29	93.83	97.46	95.42	93.93	97.87
2- MSLPQ	87.59	76.64	65.20	85.55	93.20	92.66	91.96	94.43	97.49	95.81	94.50	97.80
3- MSBSIFn	90.79	82.83	74.97	86.77	92.44	91.49	90.45	95.53	97.58	96.35	95.62	98.15
4- <i>Statistical LBP</i>												
4+1	91.59	85.09	66.96	82.80	91.36	90.27	89.02	94.53	97.64	97.58	95.62	98.02
4+2	93.40	85.79	77.36	85.77	93.47	92.45	91.29	95.93	98.19	98.05	97.10	98.60
4+3	92.58	84.88	76.15	86.02	94.08	93.10	92.06	96.25	98.09	98.11	97.25	98.42
4+3+2	93.56	85.80	77.10	86.57	94.17	93.15	92.06	96.28	98.28	98.22	97.81	98.50
4+3+2+1	93.34	86.14	74.50	86.42	93.22	92.34	91.13	95.68	98.16	98.06	97.16	98.22
3+2	91.92	86.86	75.96	88.09	93.64	92.60	91.44	96.23	98.15	98.14	97.49	98.15
Baseline [346]	/	/	/	/	/	/	/	45	67	61	54	54
[365]	/	/	/	/	/	/	/	86.5	/	/	/	95.8
[119]	/	/	/	81	/	/	/	98.5	/	/	/	99.3
[346]	/	/	/	/	/	/	/	97.4	/	/	/	98.6
[366]	/	/	/	/	/	/	/	/	/	/	95.3	97.5
[367]	/	/	/	/	/	/	/	94.10	/	/	/	96.3
[369]	/	/	/	/	/	/	/	/	82	78	72	/

Les tableaux 5.6 et 5.7 montrent que:

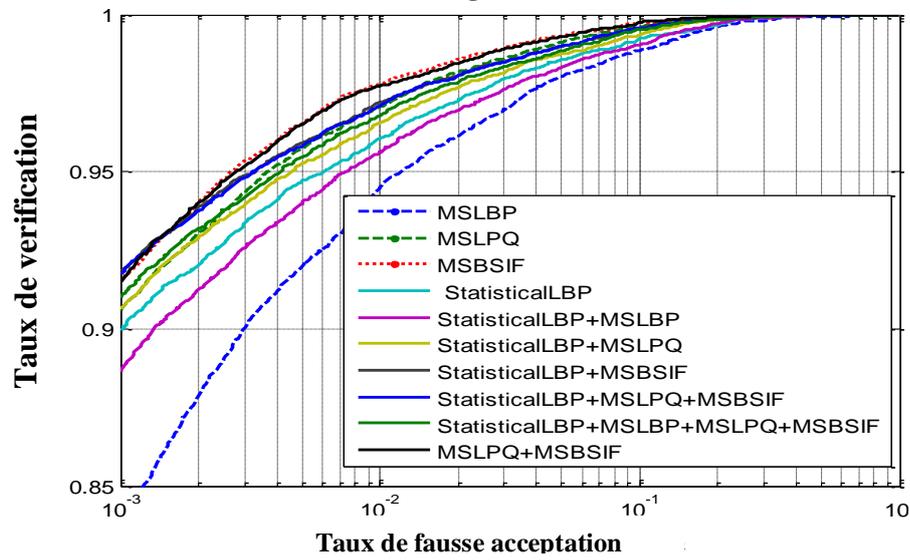
- il y a une importante amélioration de la performance pour la vérification multimodale du visage (la fusion des informations 2D et 3D) pour tous les descripteurs et multi-descripteurs dans la base de données FRGC v2,
- notre descripteur proposé *Statistical LBP* est meilleur pour la vérification du visage multimodale avec des taux de vérification de 97.58 % et 98,15 % à 0,001 TFA pour le ROC I et ROC pour le protocole "neutre contre tout", respectivement dans la base de données FRGC v2. Ce descripteur est aussi meilleur pour la base de données CASIA 3D avec TV = 97,50 %,
- la meilleure performance est obtenue par la fusion de *Statistical LBP*, MSBSIF et MSLPQ pour la CASIA 3D et les trois courbes ROC pour la FRGC v2. Mais le ROC pour le protocole "neutre contre tout" donne la meilleure performance avec la fusion de *Statistical LBP* et MSLPQ,
- sur la base FRGC v2 et pour le protocole "neutre contre tout", notre système donne les mêmes performances que celui de [369]. Il est moins bon que celui de *Mian et al.* [370]. Il est en revanche meilleur que tous les autres systèmes,
- sur la base CASIA, notre système est plus performant que les autres systèmes.

Tableau 5.7: Vérifications multi-descripteurs multimodales de visage (CASIA 3D).

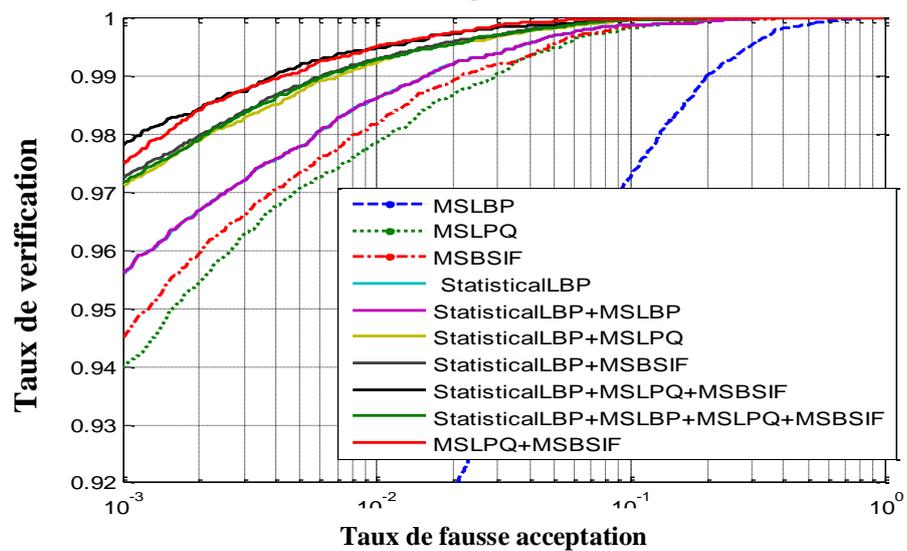
	Visage 2D		Visage 3D		Visage 2D+3D	
	TEE éval	TV test	TEE éval	TV test	TEE éval	TV test
1- MSLBP	2.61	94.10	0.80	93.60	1.18	95.69
2- MSLPQ	0.62	96.97	0.39	96.06	0.36	97.01
3- MSBSIFn	0.55	97.27	0.64	96.10	0.61	97.38
4-Statistical LBP	0.81	97.24	0.19	96.75	0.55	97.50
4+1	0.84	96.07	0.44	96.14	0.55	97.50
4+2	0.59	97.47	0.43	96.75	0.42	97.89
4+3	0.43	97.89	0.15	96.53	0.15	98.31
4+3+2	0.38	98.21	0.19	97.31	0.11	99.12
4+3+2+1	0.40	97.22	0.22	96.48	0.39	97.45
3+2	0.48	97.25	0.22	96.10	0.42	97.4
[370]	/	/	/	/	/	95
[335]	/	/	/	/	/	91.71
Notre approche chapitre 4 [371]	2.55	94.97	2.18	95.56	1.16	96.43
[341]	/	/	/	/	/	96.25



(a) Visage 2D

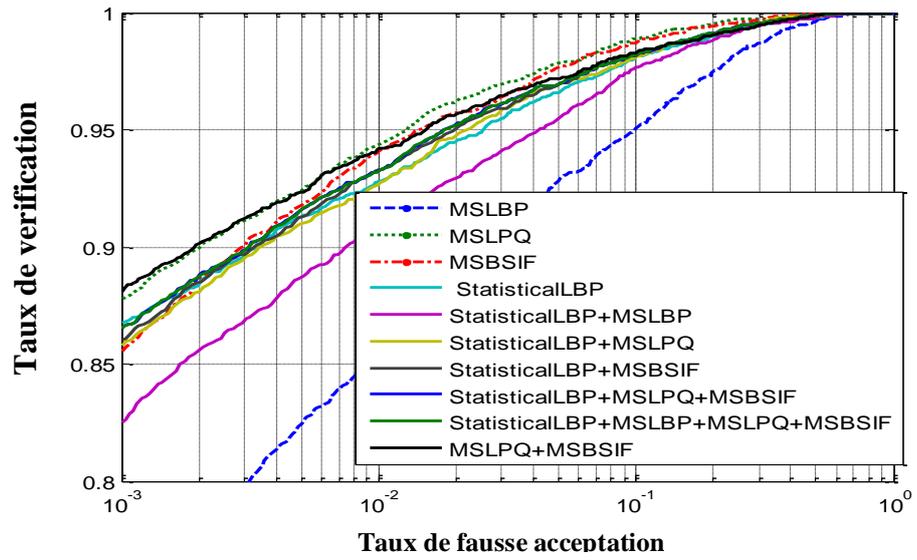


(b) Visage 3D

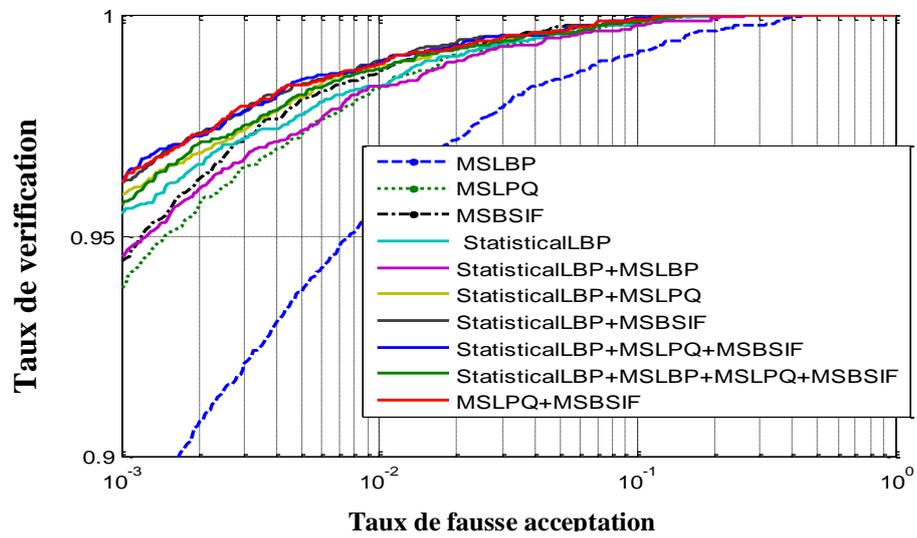


(c) Visage (2D+3D)

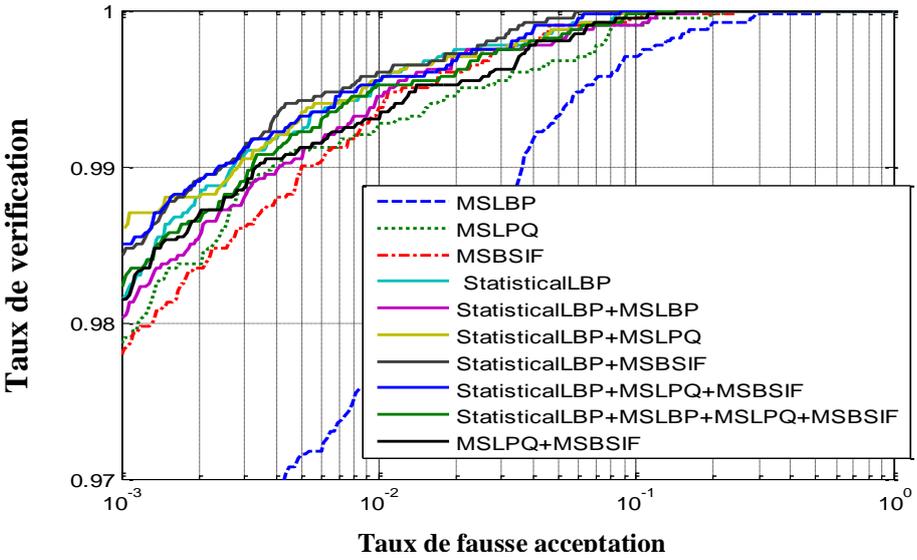
Figure 5.28: ROC III pour la vérification multi-descripteurs



(a) Visage 2D



(b) Visage 3D



(c) Visage (2D+3D)

Figure 5.29: Vérification multi-descripteurs pour le protocole "neutre contre tout"

5.5. Conclusion

Dans ce chapitre, nous avons étudié la vérification de visage uni-modale 2D et 3D et multimodale (2D+3D) basée sur la fusion de l'information locale. Notre système comprend trois étapes principales (prétraitement, extraction de caractéristiques et classification). Dans la première étape, une phase de prétraitement est nécessaire pour le remplissage de trous et le rééchantillonnage des données 2D et 3D. L'algorithme ICP (Iterative Closest Point) est utilisé pour aligner tous les visages 3D et 2D avec le premier visage 3D (expression neutre). L'approche PS (preprocessing sequence approach) est utilisée pour réduire l'influence de la variation de l'éclairage pour les images 2D. Dans la deuxième étape, nous avons utilisé quatre descripteurs locaux: LBP (Local Binary Patterns), LPQ (Local Phase Quantization), BSIF (Binarized Statistical Image Features) et *Statistical LBP* proposée). L'image faciale 2D et 3D après l'extraction de caractéristiques est divisée en 10 régions et chaque région est divisée en 15 petits blocs. Pour chacune des méthodes utilisées et pour chaque bloc, l'extraction de caractéristiques locales est résumée par les histogrammes correspondants. Ces derniers qui appartiennent à différents blocs pour la même région sont concaténés pour former un vecteur de caractéristiques représentant un descripteur spécifique à une échelle spécifique. Dans la dernière étape, nous avons proposé d'utiliser EDA (Exponential Discriminant Analysis) [19] associée à WCCN (Within Class Covariance Normalization) pour la réduction de la dimension des histogrammes. Les scores sont calculés par la métrique cosinus, normalisés par la normalisation de scores Z_{norm} et fusionnés par simple somme.

Dans ce travail, nous avons présenté quatre contributions.

- Premièrement, nous avons proposé un nouveau descripteur Statistical LBP (Statistical Local Binary Patterns) basé sur un seuillage des statistiques d'une région locale résumées par: la moyenne, la médiane, la variance, l'asymétrie et l'aplatissement avec les statistiques correspondantes des régions voisines.
- Deuxièmement, nous avons proposé un nouveau cadre pour la reconnaissance faciale basée sur l'EDA (Exponential Discriminant Analysis) associée à l'WCCN (Within Class Covariance Normalization).
- Troisièmement, nous avons proposé l'adaptation d'un descripteur BSIF sur l'image de profondeur du visage.
- Quatrièmement, un système complet d'authentification de visage (2D + 3D) est conçu avec un algorithme de prétraitement simple et des simulations ont été faites sur la texture des images 2D et la profondeur 3D et leur fusion.

Nous avons validé notre méthode proposée par comparaison avec les méthodes existantes dans l'état de l'art sur les bases de données FRGC v2 et CASIA 3D.

La reconnaissance automatique du visage a été un domaine de recherche actif au cours des quatre dernières décennies. La reconnaissance faciale a de nombreuses applications avec une importance considérable, comme l'identification biométrique, la surveillance, l'interaction homme-machine et la gestion de données multimédia. La biométrie faciale a joué un rôle important dans l'amélioration de notre sécurité en limitant la mobilité des criminels, la prévention des activités frauduleuses et la recherche des personnes disparues. Cette thèse se concentre principalement sur le thème de la reconnaissance biométrique par fusion multimodale du visage 2D et 3D.

Dans le **chapitre 1**, nous avons donné une introduction générale des contextes, les motivations, les objectifs et les contributions de cette thèse.

Dans le **chapitre 2**, nous avons passé en revue la littérature des reconnaissances faciales 2D, 3D et multimodale 2D+3D. Nous avons commencé ce chapitre par un résumé des importants algorithmes de reconnaissance de visage 2D, car la plupart des algorithmes de reconnaissance de visage 3D et 2D+3D sont des extensions directes d'algorithmes 2D existants. Le but de ce chapitre est d'évaluer les capacités et les lacunes d'algorithmes de reconnaissance de visage en cours, afin d'identifier les problèmes et les orientations pour la recherche future dans ce domaine.

De nombreuses techniques ont été proposées pour la reconnaissance de visage à l'aide des images faciales 2D. Elles peuvent être généralement divisées en deux catégories principales: les approches globales et locales. Les approches globales utilisent l'ensemble du visage comme les données d'entrée du système de reconnaissance des visages, qui sont en outre projetées sur un sous-espace de dimension relativement faible. Cela est vrai pour l'analyse de composantes principales (PCA), l'analyse discriminante linéaire (LDA), l'analyse en composantes indépendantes (ICA), etc. Alors que celles basées sur les caractéristiques locales procèdent d'abord à localiser un certain nombre de caractéristiques ou composants de l'image du visage, puis les classer en combinant et en mesurant les statistiques locales correspondantes, comme Elastic Bunch Graph Matching (EBGM), Scale-Invariant Feature Transform (SIFT) et LBP (Local Binary Patterns). Il existe aussi des méthodes qui n'appartiennent pas à l'une des deux classes, comme Actif Appearance Modèle (AAM).

Avec le développement des scanners 3D et les techniques de capture apparaissent de nombreuses approches. Semblables à la classification de reconnaissance faciale 2D, les approches de reconnaissance de visage 3D peuvent également être classées en deux approches: globales et locales. Les informations géométriques 3D, basées sur la correspondance globale, travaillent directement sur toute l'image de profondeur ou nuage de points du visage pour calculer la mesure de similarité. Les approches locales ont été largement étudiées dans la reconnaissance faciale 3D car elles procurent généralement de meilleurs résultats que les approches globales. Ces approches sont classées en cinq classes comme les méthodes basées sur le calcul de point, méthodes basées sur le calcul de courbe, méthodes basées sur les régions, méthodes basées sur les descripteurs et les méthodes basées sur la fusion des caractéristiques multiples. En plus des techniques de reconnaissance de visage 3D globales et locales, il existe aussi des approches visant à trouver une relation entre deux surfaces 3D du visage dans un espace de cartographie spéciale. Compte tenu de cela, nous les classons comme méthodes de paramétrisation de surface. Cette dernière peut être considérée comme une mise en correspondance un-à-un de la surface à un domaine approprié. En général, le domaine de paramètres sera lui-même une surface et ainsi la construction d'une cartographie des moyens de paramétrage d'une surface à une autre. Ce paramétrage peut être résumé comme un problème d'optimisation sous certaines contraintes. Les surfaces faciales 3D qui sont représentées sous forme de maillages peuvent être mappées sur une image 2D régulière de sorte que des approches plus sophistiquées de traitement du signal peuvent être appliquées.

Bien que la reconnaissance du visage 3D est vieille d'environ deux décennies, la reconnaissance multimodale de visage 2D+3D est au stade préliminaire depuis l'année 2000. Elle est très limitée et la majorité des algorithmes de cette catégorie utilisent des méthodes classiques, comme visages propres, pour créer des systèmes de reconnaissance indépendants 2D et 3D, puis de combiner les scores de

concordance par une stratégie de fusion. La troisième partie du chapitre 2 traite des modèles de reconnaissance faciale multimodales 2D+3D. Nos principaux travaux de recherche sont présentés dans les trois chapitres suivants.

Le **chapitre 3** est divisé en deux parties: vérification uni-modale et multi-algorithmes de visage 2D. Dans la première partie nous avons tout d'abord présenté notre modèle de vérification de visage. Nous avons ensuite détaillé les quatre principales étapes de ce modèle (prétraitement, extraction des caractéristiques, réduction d'espace et comparaison). En fin l'étude expérimentale de l'authentification de visages sur la base de données XM2VTS est commencée par la comparaison des algorithmes de réduction d'espace et étudié l'extraction des caractéristiques par les ondelettes de Gabor. Pour la comparaison de méthodes de réduction d'espace de données, nous avons utilisé quatre méthodes (Analyse en Composantes Principales (ACP), Analyse Discriminante Linéaire (LDA), Le modèle discriminant Linéaire amélioré de Fisher (EFM) et LDA directe (DLDA)). Dans cette étude nous déduisons que la méthode EFM est la meilleure. Pour l'extraction des caractéristiques nous avons utilisé 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 par une seuillage de la phase. Finalement, nous avons sélectionné les meilleurs systèmes d'authentification de visages pour construire un système multi-algorithmes dans la deuxième partie de ce chapitre.

En effet, dans cette dernière, les limitations des systèmes biométriques lorsqu'ils n'utilisent qu'une seule modalité biométrique sont passées en revue. Nous avons ensuite présenté une façon de réduire les limitations des systèmes biométriques monomodaux en combinant plusieurs systèmes menant ainsi à la biométrie multimodale. Les systèmes multimodaux peuvent être de différentes natures. Pour cela, nous étions amenés à étudier les différents types de combinaisons de modalités possibles, mais aussi les niveaux de fusion qui peuvent être utilisés dans un système multimodal. Nous avons présenté le type de multi-modalités et le niveau de fusion. Parmi les divers niveaux de fusion existants, nous avons choisi de travailler au niveau score car il offre le meilleur compromis entre la richesse d'information et la facilité d'implémentation. Un état de l'art de la fusion de scores, la théorie des méthodes de fusion ainsi que six méthodes de normalisation de scores utilisées sont présentées. Enfin, nous avons étudié dans les résultats expérimentaux un système multi-algorithmes d'authentification de visages. Celui-ci est construit avec les quatre systèmes retenus dans la première partie de ce chapitre. Tous d'abord nous avons présenté les distributions de scores, différentes en termes de plages de variations, dans leurs formes et leur recouvrement entre les deux classes (Client et Imposteur). Ce qui a rendu nécessaire l'étape de normalisation des scores. Nous avons ensuite comparé les méthodes de normalisation associées aux méthodes de combinaisons de scores simples. Nous pouvons dire que la somme et la somme pondérée sont les meilleures, la somme pondérée étant légèrement meilleure mais nécessitant un apprentissage des poids. Pour la combinaison de scores par logique floue à base de l'intégrale floue de Sugeno et de Choquet, nous avons obtenu une bonne performance d'authentification de visage par l'intégrale floue de Sugeno et la méthode de normalisation de scores par la fonction double sigmoïde avec un TEE=1.14% et TV=98.36%. Dans la section de comparaison des méthodes de classification de scores, nous proposons d'utiliser les méthodes de normalisation de scores avant la classification. Cela représente une contribution dans notre travail. Cette étude et les résultats obtenus nous permettent d'affirmer que les méthodes de normalisation de scores améliorent en général les performances pour toutes les méthodes de classification de scores utilisées. Les deux méthodes de classification de scores non linéaires SVM et MLP donnent des résultats équivalents, avec une simplicité d'application pour la SVM.

Tout au long du chapitre, nous avons travaillé uniquement sur la vérification de visage 2D. Malgré des résultats très intéressants, nous nous sommes rendus compte que nous atteignons des

limites en considérant une base de données contrôlée (XM2VST) qui ne contient pas de variation d'expression, changements en éclairage, rotation de la tête, bruit introduit par le capteur, ..., occlusions. Pour améliorer les performances et comme suite à notre travail nous pouvons proposer d'utiliser l'information 3D et la fusion des données 2D et 3D.

Dans le **chapitre 4**, nous avons présenté un algorithme automatique d'authentification du visage par la fusion de deux systèmes multimodaux (multi-algorithmes: la fusion de plusieurs descripteurs et multi-capteurs: la fusion de l'image de profondeur (3D) et l'intensité (2D)). Tout d'abord, nous avons corrigé la rotation de la tête par l'algorithme ICP (Iterative Closest Point), puis nous avons présenté notre algorithme de prétraitement. Pour l'extraction de caractéristiques, six méthodes d'extraction de caractéristiques locales ont été utilisées:

- descripteur MLBP (Multi-échelle Local Binary Patterns),
- caractéristiques Statistiques Locales proposées (CSL),
- ondelettes de Gabor,
- descripteur SIFT (Scale-Invariant Feature Transform),
- descripteur LPQ (Local Phase Quantization),
- descripteur BSIF (Binarized Statistical Image Features).

La classification des caractéristiques, par la métrique cosinus après réduction d'espace de données par analyse en composantes principales (ACP), suivie du modèle discriminant linéaire amélioré de Fisher (EFM). La fusion au niveau des scores par un bon classificateur à deux classes, appelé Support Vector Machines (SVM).

L'application est réalisée sur les bases de données CASIA visage 3D et Bosphorus selon notre protocole d'évaluation proposé pour traiter des problèmes majeurs dans le domaine de la reconnaissance faciale 3D et multimodale en tenant compte des variations de l'éclairage, des expressions, d'une petite rotation de la tête et des changements combinés dans l'expression sous éclairage.

Notre contribution peut se résumer comme ci-dessous.

- Étude de la fusion de deux systèmes multimodaux qui sont:
 - Multi-algorithmes: construit par la fusion de plusieurs caractéristiques locales.
 - Multi-capteur: construit par la fusion des informations 3D et 2D.
- Etude de plusieurs méthodes d'extraction de caractéristiques pour mieux comprendre leur complémentarité.
- Application pour la première fois d'un nouveau descripteur BSIF (Binarized Statistical Image Features) sur les images de profondeur (3D).
- Une nouvelle méthode d'extraction de caractéristiques (Caractéristiques Statistiques Locale (CSL)) est proposée. Elle est basée sur le calcul de paramètres statistiques dans un voisinage du pixel tel que la moyenne, l'écart type, la variance, etc.

Dans le chapitre 4, on a deux parties de résultats. En première partie, nous avons utilisé quatre descripteurs (MLBP, CSL, ondelettes de Gabor et SIFT) et les résultats sont appliqués sur une seule base de données CASIA visage 3D. Dans cette partie, nous pouvons dire que:

- PCA + EFM donne une mauvaise performance pour l'information de profondeur (3D).
- La méthode LBP pour l'information 3D donne plus de performances que l'information 2D pour les quatre valeurs de rayon (R). La fusion des quatre valeurs de rayon (MLBP) améliore les performances des informations 2D et 3D et la fusion multi-capteurs (3D + 2D) avec un TEE = 0,95% et TV = 94,54%.
- La fusion des quatre rayons ($R = 1, 2, 3, 4$) pour les différents nombres de points voisins n'améliore pas les performances de cinq descripteurs statistiques de CSL. Nous avons remarqué aussi que, pour $R = 3, R = 4$ et le nombre de points $P = 24$, on obtient un meilleur résultat pour tous les descripteurs statistiques. Par conséquent, l'augmentation du nombre de points dans le voisinage améliore le rendement dans le cas de descripteurs statistiques. Les quatre

descripteurs (*moyenne, écart-type, variance, skewness*) donnent à peu près les mêmes résultats. *Kurtosis* est le moins bon descripteur.

- La fusion des cinq paramètres de nos caractéristiques locales améliore l'authentification de visage. La performance sans *Kurtosis* est meilleure que la fusion des cinq paramètres statistiques avec TEE= 1,20% et TV = 96,32% pour la fusion multimodale (3D et 2D).
- Les meilleurs résultats par les ondelettes de Gabor sont obtenus lorsque la résolution $\lambda= 4$ (TEE = 1,57% et TV = 96,13%). La fusion de cinq résolutions n'améliore pas la performance de vérification de visages.
- La fusion de quatre LBP ($R = 1, 2, 3, 4$ et $P = 8, 16, 24$) plus SIFT donne le meilleur résultat avec un TEE = 2,48% et TV = 94. 73%.
- La fusion de notre descripteur CSL avec le descripteur MLBP + SIFT donne les meilleurs résultats avec TEE = 0,98% et TV = 97,22%.

Dans la deuxième partie de résultats, nous avons utilisé quatre descripteurs (MLBP, CSL, LPQ et BSIF). L'application est réalisée sur deux bases de données CASIA visage 3D et Bosphorus. Dans cette partie, nous pouvons déduire ce qui suit.

- Les résultats obtenus avec les deux bases de données montrent l'importance de la fusion des modalités profondeur et intensité.
- La fusion des quatre valeurs de rayon (MLBP) améliore les performances avec un TEE = 0,95% et TV = 94,54 sur la base de données CASIA; ce qui n'est pas le cas avec la base de données Bosphorus. La meilleure performance est obtenue dans le cas $(P, R) = (24,4)$ avec TEE = 4,92% et TV= 91,38.
- Nous avons obtenu une bonne performance et stabilité avec TV= 96,43% dans la base de données CASIA et TV = 97,90% dans la base de données Bosphorus pour le descripteur CSL.
- Le descripteur BSIF est performant pour des tailles du filtre supérieures à 9. La fusion des filtres est retenue pour les deux bases de données.
- LPQ donne des mauvais résultats pour R inférieur à 4 et la meilleure performance est obtenue pour la fusion des quatre grand R . Les performances d'authentification de visage sont équivalentes pour les deux bases de données.
- La fusion des trois descripteurs LBP+CSL+BSIF donne les meilleures performances avec un TV=98.41% et un TEE=1.20%, ce qui est très satisfaisant dans le cas de la base de donnée Bosphorus. On a aussi obtenu TEE=1.71% et TV=96.80% pour la base de données CASIA 3D.

Dans le **chapitre 5**, nous avons étudié la vérification de visage uni-modale 3D, puis 2D et multimodale (3D+2D) basée sur la fusion de l'information locale. Notre système comprend trois étapes principales (prétraitement, extraction de caractéristiques et classification). Dans la première étape, une phase de prétraitement est nécessaire pour remplissage de trous et l'échantillonnage des données 3D et 2D. L'algorithme ICP (Iterative Closest Point) est utilisé pour aligner tous les visages 3D et 2D avec le premier visage 3D (expression neutre) et l'approche PSA (preprocessing sequence approach) est utilisée pour réduire l'influence de la variation de l'éclairage pour les images 2D. Dans la deuxième étape nous avons utilisé quatre descripteurs locaux LBP (Local Binary Patterns), LPQ (Local Phase Quantization), BSIF (Binarized Statistical Image Features) et Statistical LBP proposée). L'image faciale 3D et 2D après extraction de caractéristiques est divisée en 10 régions et chaque région est divisée en 15 petites blocs. Pour chacune des méthodes utilisées et pour chaque bloc, l'extraction de caractéristiques locales est résumée par les histogrammes correspondants. Les histogrammes des différents blocs appartenant à la même région sont concaténés pour former un vecteur de caractéristiques représentant un descripteur spécifique à une échelle spécifique. Dans la dernière étape nous avons proposé d'utiliser EDA (Exponential Discriminant Analysis) couplé à WCCN (Within Class Covariance Normalization) pour la réduction de dimension des histogrammes. Les scores sont calculés par le métrique cosinus, normalisés par Z_norm et sont fusionnés par simple somme.

Dans ce chapitre, nous avons présenté quatre contributions. Nous avons d'abord proposé un nouveau descripteur Statistical LBP (Statistical Local Binary Patterns) basé sur un seuillage des

statistiques d'une région locale résumées par: la moyenne, la médiane, la variance, l'asymétrie et l'aplatissement avec les statistiques correspondantes des régions voisines. Deuxièmement, nous avons proposé un nouveau cadre de classification basée sur EDA (Exponential Discriminant Analysis) couplé avec WCCN (Within Class Covariance Normalization). Troisièmement, l'utilisation de stratégie multi-échelles des descripteurs locaux binaires et la proposition d'adaptation de descripteur BSIF sur l'image de profondeur de visage. Quatrièmement, un système complet d'authentification de visage 2D et 3D est conçu avec un algorithme de prétraitement simple et des simulations ont été faites pour la texture 2D et la profondeur 3D et leur fusion.

Nous avons validé l'approche proposée et l'avons comparée aux méthodes existantes dans l'état de l'art sur les bases de données FRGC v2 et CASIA 3D. En ce qui nous concerne, nous avons trouvé les résultats suivants.

- Pour les deux bases de données utilisées, nous avons obtenu à peu près les mêmes remarques et résultats.
- La stratégie multi-échelles est toujours meilleure que celle d'une seule échelle pour tous les descripteurs binaires locaux utilisés et pour la vérification uni-modale du visage 2D et 3D.
- Les descripteurs mean LBP, var LBP, skewness LBP et kurtosis LBP sont meilleurs que le descripteur LBP pour toutes les échelles.
- Le descripteur BSIF nouveau (BSIFn) est meilleur que le descripteur BSIF original (BSIFo) pour la vérification de visage 2D et 3D. MSBSIFn est le meilleur descripteur de vérification de visage 3D. Ceci nous a incité à l'utiliser comme descripteur dans les filtres intégrés par image 3D.
- Statistical LBP proposée avec $R_2 = 4$ est meilleure que multi-échelles LBP et multi-échelles LPQ pour la vérification de visage 3D. Notre descripteur statistical LBP donne la meilleure performance pour la vérification du visage 2D.
- Pour tous descripteurs binaires locaux étudiés, nous remarquons clairement qu'il y a une petite différence de performance pour les trois ROC sur la vérification faciale en 3D, mais il y a de grandes différences de performances pour la reconnaissance faciale 2D sur la base de données FRGC v2. Il y a une amélioration importante de la performance pour la vérification multimodale de visage (la fusion des informations 3D et 2D) pour tous les descripteurs et multi-descripteurs dans la base de données FRGC v2.
- Nous montrons que notre descripteur Statistical LBP est meilleur pour la vérification du visage multimodale avec des taux de vérification de 97,58% et 98,15% à 0,001 TFA pour le ROC I sur la base de données FRGC v2. Il est aussi meilleur sur la base de données CASIA 3D avec un taux de vérification TV = 97,50%.
- La meilleure performance est obtenue par la fusion de Statistical LBP et multi-échelles BSIF et multi-échelles LPQ pour la CASIA 3D et les trois courbes ROC pour la FRGC v2. Mais, pour le protocole neutre contre toute, la meilleure performance est obtenue par la fusion de Statistical LBP et multi-échelles LPQ.

Perspectives

Pour la continuité du travail présenté dans cette thèse, nous proposons ci-après trois axes de recherche.

Pour le visage 3D:

- Parce que certaines régions sont rigides par rapport aux variations d'expression alors que d'autres ne le sont pas, les différentes régions du visage sont d'une importance différente selon la précision exigée de la reconnaissance du visage. En se basant sur cette constatation, nous pouvons ainsi améliorer la performance de notre système de vérification faciale. Les impacts des différentes régions du visage sur les résultats de vérification peuvent être mesurés par des

poids tirés à partir d'une base de données spéciale qui contient de riches variations d'expression du visage, tels que les bases de données BU-3DFE et Bosphorus.

- Nous proposons d'utiliser notre descripteur Statistical LBP pour la caractérisation des images de visage 3D de basse résolution capturées avec un scanner kinect.
- Comme nous avons amélioré et adapté le descripteur BSIF dans le chapitre 5 pour l'extraction des caractéristiques du visage 3D nous proposons d'améliorer le descripteur LPQ pour les données 3D, car ce descripteur est spécialiste pour la caractérisation des images 2D floues.

Pour la Classification:

Une des conclusions de cette thèse est que la performance de la reconnaissance d'une représentation est directement liée à sa capacité de discrimination. Pour augmenter cette dernière, nous proposons:

- L'utilisation de la méthode « Sparse coding » pour la reconnaissance de visage 3D, car cette méthode reçoit un intérêt croissant pour l'apprentissage de la machine, la reconnaissance de formes, le traitement des signaux [372, 373]. Cette méthode a été appliquée avec succès à la classification d'images [374, 375] et la reconnaissance de visage [81].
- L'étude des possibilités d'accroître le caractère descriptif de la représentation en tenseur en incorporant des informations supplémentaires en elle. Un tenseur d'ordre six serait en mesure de quantifier les composantes R , V , B de la texture du visage, en plus de sa surface. Un tenseur d'ordre neuf serait capable de quantifier les coordonnées (x, y, z) , les composantes des normales à la surface et les éléments de la texture de surface R , G , B . La puissance de calcul et la mémoire seraient les facteurs limitant dans de tels cas et un équilibre entre description (précision) et efficacité doivent être pris en considération.
- L'apprentissage de la métrique (Metric Learning) a reçu beaucoup d'attention dans la vision par ordinateur et l'apprentissage de la machine au cours des dernières années. Il y a un certain nombre d'algorithmes d'apprentissage métriques dans la littérature comme: CSML (cosine similarity metric learning) [376], LMNN (large margin nearest neighbor) [377], ITML (information theoretic metric learning) [378] et NRML (neighborhood repulsed metric learning) [379]. Il serait intéressant d'étudier la fusion de ces techniques.

Pour la fusion:

- Inspiré par le fait que plusieurs longs descripteurs pourraient fournir des informations complémentaires dans la caractérisation de l'information du visage de différents points de vue et d'extraire des informations plus discriminantes, nous proposons des méthodes multi-vues d'apprentissage de la métrique comme la méthode MNRML (multiview neighborhood repulsed metric learning) [379], en vertu de laquelle plusieurs longs descripteurs peuvent être efficacement combinés et afin d'améliorer encore les performances de la vérification faciale.
- La construction d'un modèle mathématique pour la sélection ou la fusion par la sélection des meilleures caractéristiques de visage 2D et 3D tels que les modèles génétiques et les PSO (Particle Swarm Optimization) qui pourraient permettre d'optimiser notre approche.

Publications internationales

1. **Abdelmalik Ouamane**, Mébarka Belahcene, Abdelhamid Benakcha, Salah Bourennane, Abdelmalik Taleb-Ahmed, *Robust multimodal 2D and 3D face authentication using local feature fusion*, Springer: Signal, Image and Video Processing, Published online: 16 November 2014, DOI 10.1007/s11760-014-0712-x, Impact Factor =1.019.
2. **Abdelmalik Ouamane**, Mébarka Belahcene, Abdelhamid Benakcha, Mohamed Boumehrez, Abdelmalik Taleb Ahmed, *The Classification of Scores from Multi-classifiers for Face Verification*, Soft Sensors and Artificial Neural Networks, Sensors & Transducers Journal, vol. 145, no. 10, pp. 116-118, 2012.
3. **Abdelmalik Ouamane**, Mébarka Belahcene, Abdelhamid Benakcha, Mohamed Boumehrez, Abdelmalik Taleb Ahmed, *Identification of Faces by Multimodal Information Fusion of Depth and Color*, Sensors & Transducers Journal, vol. 140, no. 5, pp. 74-87, 2012.

Publications nationales

1. **Abdelmalik Ouamane**, Mébarka Belahcene, *New Multiple Classifier Systems for Face Authentication*, Courrier du Savoir, no. 18, pp.53-62, 2014.
2. Mébarka Belahcene, **Abdelmalik Ouamane**, *l'Intégrale Floue dans la Fusion d'un Système Multiclassifieurs pour la Reconnaissance de Visages*, Courrier du Savoir, no.17, pp.167-174, 2013.
3. Mébarka Belahcene, **Abdelmalik Ouamane**, Mohamed Boumehrez and Abdelmalik Benakcha, *Comparaison des Methodes de Reduction d'espace et l'application des SVMs pour la Classification dans L'authentification de Visages*, Courrier du Savoir, no. 13, pp.116-126, 2012.

Communications Internationales

1. Ammar Chouchane, Mébarka Belahcene , **Abdelmalik Ouamane**, Salah Bourennane, *A Multimodal Face Recognition Based on Histograms of Three Local Descriptors Using Score Level Fusion*, IEEE, 5th European Workshop on Visual Information Processing (EUVIP), December 2014.
2. **Abdelmalik Ouamane**, Bengherabi Messaoud, Abderrezak Guessoum, Abdenour Hadid, *Multi-Scale Multi-Descriptor Local Binary Features And Exponential discriminant Analysis For Robust Face Authentication*, IEEE, International Conference on Image Processing (ICIP), pp. 313-317, October 2014.
3. Yassine Himeur, Karima Ait-Sadi, **Abdelmalik Ouamane**, *A Fast And Robust Key-Frames Based Video Copy Detection Using Bsif-Rmi*, IEEE, 11th International Conference on Signal Processing and Multimedia Applications (SIGMAP), pp. 40-47, 2014.
4. Ammar Chouchane, Mébarka Belahcene, **Abdelmalik Ouamane**, Salah Bourennane, *3D Face Recognition Based on Histograms of Local Descriptors*, IEEE, 4th International Conference on Image Processing Theory, Tools and Applications (IPTA), October 2014.

5. Mebarka Belahcene, **Abdelmalik Ouamane**, Nawat Gaidi, *Face Identification by Fusion Scores 2D Neutral and 3D Expressions*, IEEE, 2nd International Conference on Advances in Biomedical Engineering (ICABME), pp. 113-116, September 2013.
6. **Abdelmalik Ouamane**, Mébarka Belahcène, Salah Bourennane, *Multimodal 3D and 2D Face Authentication Approach Using Extended LBP and Statistic Local Features Proposed*, IEEE, 4rd European Workshop On Visual Information Processing (EUVIP), pp. 130-135, Juin 2013.
7. **Abdelmalik Ouamane**, Mebarka Belahcene, Abdelhamid Benakcha, Abdelmalik Taleb-Ahmed, *Fusion by Classification of Scores Multi-Biometric Systems*, The Third International Conference On Image And Signal Processing And Their Applications, Mostaganem, Algeria, December 2012.
8. **Abdelmalik Ouamane**, Mebarka Belahcene, Abdelhamid Benakcha, Abderezek Benchenief, Abdelmalik Taleb Ahmed, *The Combination of Scores From a Multi-Classifiers for Face Verification*, International Conference on Embedded Systems In telecommunications and Instrumentation (ICESTI), Annaba Algeria, November 2012.
9. **Abdelmalik Ouamane**, Mebarka Belahcène, Abdelmalik Taleb-Ahmed, *Fusion d'un Système Multi-Classifieurs par l'Intégrale Floue de Sugeno et de Choquet pour la Reconnaissance de Visages*, JEAJD Polytechnique Alger, Décembre 2012.
10. Mébarka Belahcène, **Abdelmalik Ouamane**, Abdelmalik Taleb Ahmed, *Fusion By Combination Of Scores Multi-Biometric Systems*, IEEE, 3rd European Workshop on Visual Information Processing (EUVIP), pp. 252-257, 2011.
11. Mébarka Belahcene, **Abdelmalik Ouamane**, Mohamed Boumehrez et Abdelhamid Benakcha, *Authentification de Visages par les Transformations de Hough et Gabor Associées a EFM et SVM pour la Classification*, Colloque d'Informatique, Automatique et Electronique (CIAE), Université Mundiapolis, Casablanca, Maroc, 2011.

Communications nationales

1. **Abdelmalik Ouamane**, Mebarka Belahcene, Abdelhamid Benakcha, Abderezek Benchenief, Abdelmalik Taleb Ahmed, *The Classification of Scores From a Multi-Classifiers for Face Verification*, Première Conférence Nationale sur les Télécommunications (CNT) Guelma, Algerie, Novembre 2012.

- [1] K. Delac and M. Grgic. *A survey of biometric recognition methods*. 46th International Symposium Electronics in Marine, 2004.
- [2] R. Beveridge and M. Kirby. *Biometrics and Face Recognition*. IS&T Colloquium, pp. 25, 2005.
- [3] Arun A. Ross Anil K. Jain Patrick Flynn. *Handbook of biometrics*. Springer, 2007.
- [4] Karthik Nandakumar Anil K. Jain Arun A. Ross. *Introduction to biometrics*. Springer, 2011
- [5] A. Pentland and T. Choudhury. *Personalizing smart environments: Face recognition for human interaction*. IEEE Computer. Special issue on Biometrics, 2000.
- [6] H. LI. *Towards Three-Dimensional Face Recognition in the Real*, PhD thesis L'ECOLE CENTRALE DE LYON, 18 November 2013.
- [7] R. Hietmeyer. *Biometric Identification promises fast and secure processing of airline passengers*. International Civil Aviation Organization Journal, vol. 55, no. 9, pp. 10-11, 2000.
- [8] P. Buysens, *Fusion de différents modes de capture pour la reconnaissance du visage appliquée aux e_transactions*, PhD thesis l'université de caen spécialité : informatique et applications, 07/07/2006
- [9] S. Z. Li and A. K. Jain. *Handbook of face recognition*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
- [10] http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html
- [11] <http://www.lrv.fri.uni-lj.si/facedb.html>
- [12] <http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html>
- [13] <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>
- [14] http://www.ri.cmu.edu/projects/project_418.html
- [15] <http://images.ee.umist.ac.uk/danny/database.html>
- [16] <http://www.uk.research.att.com/facedatabase.html>
- [17] <http://www.mis.atr.co.jp/~mlyons/jaffe.html>
- [18] <http://www.humanscan.de/support/downloads/facedb.php>
- [19] <http://www.ee.oulu.fi/research/imag/color/pbfd.html>
- [20] <http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/>
- [21] <http://www.itl.nist.gov/iad/humanid/feret/>
- [22] <http://vis-www.cs.umass.edu/lfw/>
- [23] T. C. Faltemier. *Flexible and robust 3D face recognition*. PhD thesis, The University of Notre Dame, 2007.
- [24] P. Sinha, B. J. Balas, Y. Ostrovsky and R. Russell. *Face recognition by humans. Face Processing: Advanced Modeling and Methods*. Academic Press, San Diego, 2006.
- [25] K. W. Bowyer, K. Chang and P. Flynn. *A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition*. Computer Vision and Image Understanding, vol. 101, pp. 1-15, 2006.
- [26] A. S. Mian, M. Bennamoun and R. A. Owens. *An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 11, pp. 1927-1943, 2007.
- [27] C. Beumier and M. Acheroy. *3D Facial Surface Acquisition by Structured Light*. In In International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging, pp. 103-106, 1999.
- [28] P. S Huang, C. Zhang and F. P. Chiang. *High-speed 3-D shape measurement based on digital fringe projection*. Optical Engineering, vol. 42, no. 1, pp. 163-168, 2003.
- [29] S. Zhang and P. Huang. *High-Resolution, Real-time 3D Shape Acquisition*. In Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop, CVPRW '04, 2004.
- [30] F. Tsalakanidou, F. Forster, S. Malassiotis and M. G. Strintzis. *Real-time acquisition of depth and color images using structured light and its application to 3D face recognition*. Real-Time Imaging, vol. 11, no. 5-6, pp. 358-369, 2005.

- [31] S. Zhang and S. T. Yau. *High-resolution, real-time 3D absolute coordinate measurement based on a phase-shifting method*. Optics Express, vol. 14, no. 7, pp. 2644-2649, 2006.
- [32] Minoltavid910url.<http://www.konicaminolta.com/instruments/products/3d/noncontact/vivid910/features.html>, 2011.
- [33] Inspeck mega capturo ii digitizerurl. <http://www.inspeck.com>, 2011.
- [34] Kinecturl. <http://www.xbox.com/en-GB/kinect>, 2011.
- [35] F. Tsalakanidou, F. Forster, S. Malassiotis and M. G. Strintzis. *Real-time acquisition of depth and color images using structured light and its application to 3D face recognition*. Real-Time Imaging, vol. 11, no. 5-6, pp. 358-369, 2005.
- [36] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski. *A comparison and evaluation of multi-view stereo reconstruction algorithms*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 519-528, 2006.
- [37] Di4d: 4d capture system url. <http://www.di3d.com/products/4dsystems>, 2011
- [38] 3dmd: 4d capture system url. <http://www.di3d.com/products/4dsystems>, 2011.
- [39] L. Yin, X. Chen, Y. Sun, T. Worm and M. Reale. *A high-resolution 3D dynamic facial expression database*. In 8th IEEE International Conference on Automatic Face and Gesture Recognition, FG'08., pp. 1-6, 2008.
- [40] L. Benedikt, D. Cosker, P. L. Rosin and D. Marshall. *Assessing the uniqueness and permanence of facial actions for use in biometric applications*. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 40, no. 3, pp. 449-460, 2010.
- [41] D. Cosker, E. Krumhuber and A. Hilton. *A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling*. In 2011 IEEE International Conference on Computer Vision (ICCV), pp. 2296-2303. IEEE, 2011.
- [42] T. Beeler, B. Bickel, P. Beardsley, B. Sumner and M. Gross. *High-quality single-shot capture of facial geometry*. ACM Transactions on Graphics (TOG), vol. 29, no. 4, pp. 40, 2010.
- [43] R. J. Woodham. *Photometric method for determining surface orientation from multiple images*. Optical Engineering, vol. 19, no. 1, pp. 191139-191139, 1980.
- [44] G. Brostow, C. Hernandez, G. Vogiatzis, B. Stenger, R. Cipolla, *Video normals from colored lights*, IEEE Trans. Pattern Anal. Mach. Intell. 1 (2011).
- [45] M. F. Hansen, G. A. Atkinson, L. N. Smith and M. L. Smith. *3D face reconstructions from photometric stereo using near infrared and visible light*. Computer Vision and Image Understanding, vol. 114, no. 8, pp. 942-951, 2010.
- [46] Robert T. Frankot, R. Chellappa and S. Member. *A Method for enforcing integrability in shape from shading algorithms*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 10, pp. 439-451, 1988.
- [47] S. Zafeiriou, M. Hansen, G. Atkinson, V. Argyriou, M. Petrou, M. Smith and L. Smith. *The photoface database*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Workshop on Biometrics) (CVPRW2011), 2011.
- [48] K. Messer, J. Kittler, M. Sadeghi, S. Marcel, C. Marcel, S. Bengio, F. Cardinaux, C. Sanderson, J. Czyz, L. Vandendorpe, S. Srisuk, M. Petrou, W. Kurutach, A. Kadyrov, R. Paredes, E. Kadyrov, B. Kepenekci, F. B. Tek, G. B. Akar, N. Mavity and F. Deravi. *Face Verification Competition on the XM2VTS Database*. In International Conference on Audio and Video Based Biometric Person Authentication, pp. 964-974, 2003
- [49] C. Beumier and M. Acheroy. *Face verification from 3D and grey level clues*. Pattern Recognition Letters, vol. 22, no. 12, pp. 1321-1329, 2001
- [50] N. F. Troje and H. H. Bultho. *Face recognition under varying poses: the role of texture and shape*. Vision Research, vol. 36, no. 12, pp. 1761-1771, 1996
- [51] T. Heseltine. *Face recognition: two-dimensional and three-dimensional techniques*. PhD thesis, The University of York, 2005

- [52] A. B. Moreno and A. Sanchez. *GavabDB: A 3D Face Database*. In COST Workshop on Biometrics on the Internet: Fundamentals, Advances and Applications, pp. 77-82, 2004
- [53] L. Yin, X. Wei, Y. Sun, J. Wang and M. J. Rosato. *A 3D facial expression database for facial behavior research*. In International Conference on Automatic Face and Gesture Recognition, pp. 211-216, 2006
- [54] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur and L. Akarun. *Bosphorus database for 3D face analysis*. In COST Workshop on Biometrics and Identity Management, 2008
- [55] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, *Overview of the face recognition grand challenge*. in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1. Washington, DC: IEEE Computer Society, 2005, pp. 947-954.
- [56] C. Conde, A. Serrano, L. J. Rodriguez-Aragon and E. Cabello. *An automatic 2D, 2.5D & 3D score-based fusion face verification system*. In IEEE International Conference on Application-Specific Systems, Architectures and Processors, pp. 214-219, 2007
- [57] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In Annual Conference on Computer Graphics, pp. 187-194, 1999.
- [58] Multimedia and Intelligent Software Technology Beijing Municipal Key Laboratory. *The BJUT-3D large-scale chinese face database*. Beijing Univ. Technology, Beijing, China, Tech. Rep. MISKL-TR-05-FMFR-001, Aug. 2005.
- [59] K. I. Chang, K. W. Bowyer, and P. J. Flynn. *Face recognition using 2D and 3D facial data*. in Proc. ACM Workshop Multimodal User Authentication, pp. 25-32, Dec. 2003.
- [60] P. J. Flynn, K. W. Bowyer, and P. J. Phillips. *Assessment of time dependency in face recognition: An initial study*. in Proc. 4th Int. Conf. Audio Video-Based Biometric Person Authentication (AVBPA), LNCS 2688, pp. 44-51, 2003.
- [61] C. Xu, T. Tan, S. Z. Li, Y. Wang, and C. Zhong. *Learning effective intrinsic features to boost 3D-based face recognition*. in Proc. 9th Eur. Conf. Comput. Vis. (ECCV), Part II, LNCS 3952, pp. 416-427, May 2006.
- [62] . Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik. *Texas 3D face recognition database*. in Proc. Southwest Symp. Image Anal. Interpret. (SSIAI), pp. 97-100.
- [63] X. Lu and A. K. Jain. *Deformation modeling for robust 3D face matching*. in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Washington, DC: IEEE Computer Society, pp. 1377-1383, Jun. 2006.
- [64] Y. Wang, G. Pan, Z. Wu, and Y. Wang. *Exploring facial expression effects in 3D face recognition using partial ICP*. in Proc. Asian Conf. Comput. Vision (Lectures Notes in Computer Science 3851), P. Narayanan, Ed. Berlin, Germany: Springer-Verlag, pp. 581-590, Jan. 2006.
- [65] M. Daoudi, F. ter Haar, and R. C. Veltkamp. *Shape retrieval contest of 3D face scans*. SHREC, 2008.
- [66] D. Smeets, P. Claes, J. Hermans, D. Vandermeulen, and P. Suetens. *Comparative Study of 3-D Face Recognition Under Expression Variations*. IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications And Reviews, vol. 42, no. 5, SEPTEMBER 2012.
- [67] WW Bledsoe. *Man-machine facial recognition*. Panoramic Research Inc., Palo Alto, CA, 1966.
- [68] M. H. Mahoor and M. Abdel-Mottaleb. *3D face recognition based on 3D ridge lines in range data*. In IEEE International Conference on Image Processing, vol. 1, pp. 137-140, 2007.
- [69] P.J. Phillips, P.J. Grother, R.J. Micheals, D.M. Blackburn, E. Tabassi, and J.M. Bone. *Face recognition vendor test 2002: overview and summary*. Technical report, NISTIR 6965, National Institute of Standards and Technology, 2003.
- [70] J. Zou, Q. Ji and G. Nagy. *A comparative study of local matching approach for face recognition*. IEEE Transactions on Image Processing, vol. 16, no. 10, pp. 2617-2628, 2007.

- [71] W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld. *Face recognition: A literature survey*. ACM Computing Survey, vol. 35, pp. 399-458, Dec. 2003.
- [72] M. Turk and A. Pentland. *Eigenfaces for face recognition*. Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.
- [73] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman. *Eigenfaces vs. Fisherfaces: recognition using class specific linear projection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, 1997.
- [74] L. Wiskott, J. M. Fellous, N. Kuiger, and C. von der Malsburg. *Face recognition by elastic bunch graph matching*. IEEE Trans. PAMI, vol. 19, no. 7, pp. 775-779, July 1997.
- [75] M. Bicego, A. Lagorio, E. Grosso and M. Tistarelli. *On the use of SIFT features for face authentication*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop, pp. 35-40, 2006.
- [76] T. Ahonen, A. Hadid and M. Pietikainen. *Face Description with Local Binary Patterns: Application to Face Recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037-2041, 2006.
- [77] I. L. Dryden, Kanti V. Mardia, *Statistical Shape Analysis*, Wiley, 1998.
- [78] N. Morizet. *Reconnaissance Biométrique par Fusion Multimodale du Visage et de l'Iris*. Thèse de l'Ecole Nationale Supérieure des Télécommunications Spécialité : Signal et Images, 2009.
- [79] M. A. Turk and A. P. Pentland. *Face recognition using eigenfaces*. In IEEE Conference on Computer Vision and Pattern Recognition, pp. 586-590, June 1992.
- [80] K. Pearson. On lines and planes of closest fit to points in space. Philosophical Magazine, vol. 2, pp. 559-572, 1901.
- [81] D. Huang, *Robust Face Recognition based on Three Dimensional Tata*, Thèse de l'école central de LYON spécialité: Informatique, 2011.
- [82] Y. Jian, D. Zhang, A. Frangi, and J.-Y. Yang. *Two-Dimensional PCA : A New Approach to Appearance-Based Face Representation and Recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 1, pp. 131-137, January 2004.
- [83] B. Ans, J. Héroult, and C. Jutten. *Adaptive neural architectures : detection of primitives*. In COGNITIVA, pp. 593-597, 1985.
- [84] J. Héroult, B. Ans, and C. Jutten. *Circuits neuronaux à synapses modifiables : Décodage de messages composites par apprentissage non supervisé*. In Comptes Rendus de l'Académie des Sciences, pp. 525-528, 1984.
- [85] P. Comon. *Independent component analysis - a new concept ?*. Signal Processing, vol. 36, no. 3, pp. 287-314, 1994.
- [86] C. Jutten and J. Héroult. *Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture*. Signal Processing, vol. 24, pp. 1-10, 1991.
- [87] J. F. Cardoso. *Infomax and maximum likelihood for blind source separation*. IEEE Signal Processing Letters, March 23 1999.
- [88] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent component analysis*. John Wiley & Sons, pp. 1-12, 2001.
- [89] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. *Face recognition by independent component analysis*. Transactions on Neural Networks, August 15 2002.
- [90] B. A. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridge. *Recognizing faces with PCA and ICA*. Computer Vision and Image Understanding, vol. 91, no. 1-2, pp. 115-137, July/August 2003.
- [91] W. S. Torgerson : *Multidimensional scaling*. Psychometrika, vol. 17, no. 401-419, 1952.
- [92] T. Cox et M. Cox. *Multidimensional scaling*. Chapman and Hall, 2000.
- [93] C. K. I. Williams, *On a connection between kernel pca and metric multidimensional scaling*. Machine Learning, vol. 46, no. 1-3, pp. 11-19, 2002.
- [94] D. D. Lee and H. S. Seung. *Learning the parts of objects by non-negative matrix factorization*. Nature, vol. 401, pp. 788-791, October 1999.

- [95] P. Belhumeur, J. Hespanha, and D. Kriegman. *Eigenfaces vs. Fisherfaces : Recognition Using Class Specific Linear Projection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, pp. 711-720, 1997.
- [96] R. A Fisher. *The use of multiple measures in taxonomic problems*. Ann. Eugenics, 7:179-188, 1936.
- [97] Daniel L. Swets and J. Weng. *Using discriminant eigenfeatures for image retrieval*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18no. 8, pp. 831-836, 1996.
- [98] A. Pentland, B. Moghaddam and T. Starner. *View-based and modular eigenspaces for face recognition*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994.
- [99] B. Moghaddam. *Principal manifolds and probabilistic subspaces for Visual Recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 6, pp. 780-788, 2002.
- [100] J. Lu, K. N. Plataniotis and A. N. Venetsanopoulos. *Face recognition using LDA-based algorithms*. IEEE Transactions on Neural Networks, vol. 14, pp. 195-200, 2003.
- [101] M. A. O. Vasilescu and D. Terzopoulos. *Multilinear analysis of image ensembles: Tensor Faces*. In European Conference on Computer Vision, pp. 447-460, 2002.
- [102] J. H. Friedman. *Regularized discriminant analysis*. J. Amer. Stat. Assoc., vol. 84, no. 405, pp. 165-175, 1989.
- [103] T. Hastie, A. Buja, and R. Tibshirani. *Penalized discriminant analysis*. Ann. Stat., vol. 23, no. 1, pp. 73-102, 1995.
- [104] L. Chen, H. Liao, M. Ko, L. J., and G. Yu. *A New LDA-based Face Recognition System Which Can Solve the Small Samples Size Problem*. Journal of Pattern Recognition, vol.33, pp.1713-1726, 2000.
- [105] W. Liu, Y. Wang, S. Z. Li, and T. Tan. *Null Space Approach of Fisher Discriminant Analysis for Face Recognition*. in *Biometric Authentication*, Lecture Notes in Computer Science: Springer , vol.3087, pp.32-44, 2004.
- [106] X. Wang and X. Tang. *Dual-space Linear Discriminant Analysis for Face Recognition*. in Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp.564-566, 2004.
- [107] M. Loog, R. P. W. Duin, and R. Haeb-Umbach. *Multiclass Linear Dimension Reduction by Weighted Pairwise Fisher Criteria*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.23, pp.762-766, 2001.
- [108] P. Howland and H. Park. *Generalized Discriminant Analysis Using the Generalized Singular Value Decomposition*. IEEE Trans. On Pattern Analysis and Machine Intelligence, vol.26, pp.995-1006, 2004.
- [109] J. P. Ye, R. Janardan, C. H. Park, and H. Park. *An Optimization Criterion for Generalized Discriminant Analysis on Undersampled Problems*. IEEE Trans. On Pattern Analysis and Machine Intelligence, vol.26, pp. 982-994, 2004.
- [110] J. W. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. *Boosting Linear Discriminant Analysis for Face Recognition*. in Proceedings of IEEE International Conference on Image Processing, vol.1, pp. 657-660, 2003.
- [111] Q. Yang and X. Q. Ding. *Discriminant Local Feature Analysis of Facial Images*. in IEEE International Conference on Image Processing, vol.2, pp. 863-866, 2003.
- [112] V. D. M. Nhat and S. Lee. *Block LDA for Face Recognition*. in Computational Intelligence and Bioinspired Systems, Lecture Notes in Computer Science: Springer Berlin / Heidelberg, vol. 3512, pp. 899-905, 2005.
- [113] D. Zhou and X. Yang. *Face Recognition Using Enhanced Fisher Linear Discriminant Model with Facial Combined Feature*. in PRICAI 2004: Trends in Artificial Intelligence, vol.3157, pp. 769-777, 2004.

- [114] W. C. Zhang, S. G. Shan, W. Gao, Y. Z. Chang, and B. Cao. *Component-based cascade linear discriminant analysis for face recognition*. in Advances In Biometric Person Authentication, Proceedings, vol.3338, pp. 288-295, 2004.
- [115] H. Cevikalp, M. Neamtu, M. Wilkes, A. Barkana. *Discriminative common vectors for face recognition*. IEEE Trans. Pattern Anal. Machine Intell. vol. 27, no. 1, pp. 4-13, 2005.
- [116] M. Visani, C. Garcia, and J. M. Jolion. *Normalized radial basis function networks and bilinear discriminant analysis for face recognition*. In IEEE Conference on Advanced Video and Signal Based Surveillance, pp. 342-347, 2005.
- [117] B. Scholkopf, A. Smola, and K. Muller. *Nonlinear component analysis as a kernel eigenvalue problem*. Neural Computation, pp. 1299-1319, 1998.
- [118] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. CUP, June 2004.
- [119] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, G. Medioni, *Performance of Geometrix Active IDTM 3D Face Recognition Engine on the FRGC Data*, in CVPR 05 : the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, pp. 154. 2005.
- [120] H. Hoffmann. *Kernel PCA for novelty detection*. Pattern Recognition, vol. 40, no. 3, pp. 863-874, March 2007.
- [121] F. Bach and M. Jordan. *Kernel independent component analysis*. Journal of Machine Learning Research, pp. 1-48, 2002.
- [122] K. Q. Weinberger, and L. K. Saul. *Unsupervised learning of image manifolds by semidefinite programming*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR-04), vol 2, pp. 988-995, 2004.
- [123] M. H. Yang. *Face recognition using extended isomap*. In International Conference on Image Processing, pp. 117-120, 2002.
- [124] G. Hagen, T. Smith, A. Banasuk, R.R. Coifman, and I. Mezic. *Validation of low-dimensional models using diffusion maps and harmonic averaging*. In IEEE Conference on Decision and Control, 2007.
- [125] J. Wang, C. Zhang, and Z. Kou. *An analytical mapping for LLE and its application in multi-pose face synthesis*. In British Machine Vision Conference, 2003.
- [126] D. A. Socolinsky and A. Selinger. *Thermal face recognition in an operational scenario*. In IEEE Conference on Computer Vision and Pattern Recognition, pp. 1012-1019, 2004.
- [127] X. He, S. C. Yan, Y. X. Hu, and H. J. Zhang. *Learning a Locality Preserving Subspace for Visual Recognition*. in Proceedings of 9th IEEE International Conference on Computer Vision, vol. 1, pp. 385-392, 2003.
- [128] S. C. Yan, H. J. Zhang, Y. X. Hu, B. Y. Zhang, and Q. S. Cheng. *Discriminant Analysis on Embedded Manifold*. in European Conference on Computer Vision, vol. 3021, pp.121-132, 2004.
- [129] J. Zhang, S. Z. Li, and J. Wang. *Nearest Manifold Approach for Face Recognition*. in Proc. IEEE International Conference on Automatic Face and Gesture Recognition, pp.223-228, 2004.
- [130] Y. Wu, K. L. Chan, and L. Wang. *Face Recognition based on Discriminative Manifold Learning*. in Proc. IEEE Int'l Conf. on Pattern Recognition, vol.4, pp.171-174, 2004.
- [131] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. *Face recognition using laplacianfaces*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pp. 328-340, 2005.
- [132] S. Harmeling. *Independent component analysis and beyond*. PhD thesis, Universität Potsdam ; Mathematisch-Naturwissenschaftliche Fakultät. Institut für Informatik, 2004.
- [133] B. Raducanu and F. Dornaika. *Dynamic facial expression recognition using laplacian eigenmaps-based manifold learning*. In International Conference on Robotics and Automation, pp. 156-161, 2010.
- [134] H. Kim, H. Park, and H. Zha. *Distance preserving dimension reduction for manifold learning*. In International Conference on Data Mining. SIAM, 2007.
- [135] Q. Wang and J. Li. *Combining local and global information for nonlinear dimensionality reduction*. Neurocomputing, vol. 72, pp. 2235-2241, 2009.

- [136] M. K. Fleming and G. W. Cottrell. *Categorization of faces using unsupervised feature extraction*. In IEEE International Joint Conference on Neural Networks, volume II, pp. II-65-II-70, 1990.
- [137] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back. *Face recognition : A convolutional neural-network approach*. IEEE Transactions on Neural Networks, vol. 8, no. 1, pp. 98-113, January 1997.
- [138] S. Duffner and C. Garcia. *Face recognition using non-linear image reconstruction*. In i-LIDS : Bag and Vehicle Detection Challenge, pp. 459-464, 2007.
- [139] T. Zhang, B. Fang, Y. Y. Tang, Z. W. Shang, and B. Xu. *Generalized discriminant analysis: A matrix exponential approach*. IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 1, pp. 186-197, 2010.
- [140] Ngoc Son VU. *Contributions à la reconnaissance de visages à partir d'une seule image et dans un contexte non-contrôlé*, Thèse de L'Université de Grenoble, 2010.
- [141] T. Kanade. *Computer recognition of Human faces*. 1977.
- [142] B. S. Manjunath, R. Chellappa, and C. von der Malsburg. *A feature based approach to face recognition*. In CVPR, 1992.
- [143] R. Brunelli and T. Poggio. *Face recognition : features versus templates*. IEEE Trans. PAMI, vol. 15, no. 10, pp. 1042-1052, 1993.
- [144] H. A. Rowley, S. Baluja, and T. Kanade. *Neural network-based face detection*. IEEE Trans. PAMI, vol. 20, no. 1, pp. 23-38, 1998.

- [145] A. Lanitis. *Automatic face identification system using exible appearance models*. Image and Vision Computing, vol. 13, no. 5, pp. 393-401, 1995.
- [146] 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*. IEEE Trans. on Computers, vol. 42, no. 3, pp. 300-311, 1993.
- [147] B. Duc, S. Fischer, and J. Bigun. *Face authentication with gabor information on deformable graphs*. IEEE Trans. on Image Processing, 1999.
- [148] Y. Gao and Y. Qi. *Robust visual similarity retrieval in single model face databases*. Pattern Recognition, vol. 38, no. 7, pp. 1009-020, 2005.
- [149] P. Viola and M.J. Jones. *Robust real-time face detection*. Int Journal of Computer Vision, vol. 57, pp. 137-154, 2004.
- [150] D.G. Lowe. *Distinctive image features from scale-invariant keypoints*. Int Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.
- [151] T. Ahonen, A. Hadid, and M. Pietikainen. *Face recognition with local binary patterns*. In ECCV, pp. 469-481, 2004.
- [152] V. Ojansivu and J. Heikkila. *Blur insensitive texture classification using local phase quantization*. International Conference on Image and Signal Processing (ICISP08), pp. 236-243, 2008.
- [153] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, W. Gao, *WLD: A Robust Local Image Descriptor*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32 , no. 9, 2010.
- [154] J. Kannala and E. Rahtu. *BSIF: Binarized Statistical Image Features*. 21st International Conference on Pattern Recognition (ICPR), pp.1363-1366, 2012.
- [155] D. Huang, Caifeng, M. Ardabilian, Y. Wang, and L. Chen. *Local Binary Patterns and Its Application to Facial Image Analysis: A Survey*, IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 41, no. 6, november 2011.
- [156] T. Ojala, M. Pietikäinen, and T. Maenpaa. *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 97-987, Jul. 2002.

- [157] H. Jin, Q. Liu, H. Lu, and X. Tong. *Face detection using improved LBP under Bayesian framework*. in Proc Int. Conf. Image Graph., 2004, pp. 306-309.
- [158] H. Jin, Q. Liu, X. Tang, and H. Lu. *Learning local descriptors for face detection*. in Proc. IEEE Int. Conf. Multimedia Expo., pp. 928-931, Jul. 2005.
- [159] B. Froba and A. Ernst. *Face detection with the modified census transform*. in Proc. IEEE Int. Conf. Autom. Face Gesture Recog., pp. 91-96, 2004.
- [160] J. Ruiz-del-Solar and J. Quinteros. *Illumination compensation and normalization in eigenspace-based face recognition: A comparative study of different pre-processing approaches*. Pattern Recog. Lett., vol. 29, no. 14, pp. 1966-1979, 2008.
- [161] G. Bai, Y. Zhu, and Z. Ding. *A hierarchical face recognition method based on local binary pattern*. in Proc. Congr. Image Signal Process., pp. II: 610-614, May 2008.
- [162] S. Marcel, Y. Rodriguez, and G. Heusch. *On the recent use of local binary patterns for face authentication*. Dalle Molle Inst. Perceptual Artif. Intell. (IDIAP) Res. Inst., Martigny, Switzerland, Res. Rep. IDIAPRR, pp. 34-2006, 2006.
- [163] H. Yang and Y. Wang. *A LBP-based face recognition method with Hamming distance constraint*. in Proc. Int. Conf. Image Graph., pp. 645-649, Aug. 2007.
- [164] D. Huang, Y. Wang, and Y. Wang. *A robust method for near infrared face recognition based on extended local binary pattern*. in Proc. Int. Symp. Vis. Comput., pp. 437-446, 2007.
- [165] Y. Huang, Y. Wang, and T. Tan. *Combining statistics of geometrical and correlative features for 3D face recognition*. in Proc. Brit. Mach. Vis. Conf., pp. III: 879_888, 2006.
- [166] Z. Guo, L. Zhang, and D. Zhang. *A completed modeling of local binary pattern operator for texture classification*. IEEE Trans. Image Process., vol. 19, no. 6, pp. 1657-1663, Jun. 2010.
- [167] X. Tan and B. Triggs. *Enhanced local texture feature sets for face recognition under difficult lighting conditions*. in Proc. Anal. Model. Faces Gestures, pp. 168-182, 2007.
- [168] T. Ahonen and M. Pietikäinen. *Soft histograms for local binary patterns*. in Proc. Fin. Signal Process. Symp., Oulu, Finland, 2007.
- [169] S. Liao and A. C. S. Chung. *Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude*. in Proc. Asian Conf. Comput. Vis., pp. 672-679, 2007.
- [170] L. Zhang, R. Chu, S. Xiang, and S. Z. Li, *Face detection based on Multi-Block LBP representation*. in Proc. Int. Conf. Biometrics, pp. 11-18, 2007.
- [171] S. Liao and S. Z. Li. *Learning multi-scale block local binary patterns for face recognition*. in Proc. Int. Conf. Biometrics, pp. 828-837, 2007.
- [172] L. Wolf, T. Hassner, and Y. Taigman. *Descriptor based methods in the wild*. in Proc. ECCV Workshop Faces 'Real-Life' Images: Detection, Alignment, Recog., Marseille, France, 2008.
- [173] G. J. Edwards, T. F. Cootes and C. J. Taylor. *Face recognition using active appearance models*. In European Conference on Computer Vision, pp. 581-595, 1998.
- [174] F. Cardinaux, C. Sanderson and S. Marcel. *Comparison of MLP and GMM classifiers for face verification on XM2VTS*. In International on Audio and Video Based Biometric Person Authentication, pp. 964-974, 2003.
- [175] S. Lucey and T. Chen. *A GMM parts based face representation for improved verification through relevance adaptation*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 855-861, 2004.
- [176] F. Samaria and A. Harter. *Parameterisation of a stochastic model for human face identification*. In Workshop on Applications of Computer Vision, 1994.
- [177] A. V. Nefian and M. H. Hayes II I. *Face Recognition Using An Embedded HMM*. In International Conference on Audio- and Video-based Biometric Person Authentication, 1999.
- [178] T.F. Cootes and G.J. Edwards and C.J. Taylor. *Active Appearance Models*. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 681-685, 2001.
- [179] I. L. Dryden, Kanti V. Mardia, *Statistical Shape Analysis*, Wiley, 1998.

- [180] T.F. Cootes and P. Kittipanya-ngam, *Comparing Variations on the Active Appearance Model Algorithm*. British Machine Vision Conference, pp. 837-846, 2002.
- [181] B. Heisel, P. Ho, J. Wu and T. Poggio. *Face recognition component-based versus global approaches*. Computer Vision and Image Understanding, vol. 91, no. 1-2, pp.6-21, 2003.
- [182] X. Tan, S. Chen, Z. Zhou and F. Zhang. *Face recognition from a single image per person: a survey*. Pattern Recognition, vol. 39, no. 9, pp. 1725-1745, 2006.
- [183] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade and B. Lu. *Person specific SIFT features for face recognition*. In IEEE International Conference on Acoustics, Speech, and Signal Processing, 2007.
- [184] J. Ruiz del Solar, R. Verschae and M. Correa. *Recognition of faces in unconstrained environments: a comparative study*. Eurasip Journal on Advances in Signal Processing, vol. 2009, no. 1, pp. 1-20, 2009.
- [185] B. Achermann, X. Jiang and H. Bunke. *Face recognition using range images*. In International Conference on Virtual Systems and MultiMedia, pp. 129-136, 1997.
- [186] C. Heshner, A. Srivastava and G. Erlebacher. *A novel technique for face recognition using range imaging*. In International Symposium on Signal Processing and Its Applications, vol. 2, pp. 201-204, Jul. 2003.
- [187] T. Heseltine, N. Pears and J. Austin. *Three-dimensional face recognition: an eigensurface approach*. In IEEE International Conference on Image Processing, vol. 2, pp. 1421-1424, Oct. 2004.
- [188] T. Heseltine, N. Pears and J. Austin. *Three-dimensional face recognition: a fishersurface approach*. In International Conference on Image Analysis and Recognition, pp. 684-691, 2004.
- [189] P. J. Besl and H. D. McKay. *A method for registration of 3-D shapes*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pp. 239-256, Feb. 1992.
- [190] G. Medioni and R. Waupotitsch. *Face Modeling and Recognition in 3-D*. In IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp. 232-233, 2003.
- [191] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis and T. Theoharis. *Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 4, pp. 640-649, 2007.
- [192] X. Li, T. Jia and H. Zhang. *Expression-insensitive 3D face recognition using sparse representation*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2575-2582, Jun. 2009.
- [193] H. Mohammadzade and D. Hatzinakos, *Iterative Closest Normal Point for 3D Face Recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 2, 2013.
- [194] M. Ardabilian, *Contributions à la modélisation et à la reconnaissance faciales 3D*, THESE pour obtenir le grade de Docteur en Informatique de l'Ecole Centrale de Lyon, 2006.
- [195] D. P. Huttenlocher, G. A. Klanderman and W. J. Ruck-lidge. *Comparing images using the Hausdorff distance*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 9, pp. 850-863, 1993.
- [196] B. Achermann and H. Bunke. *Classifying range images of human faces with Hausdorff distance*. In International Conference on Pattern Recognition, vol. 2, pp. 809-813, 2000.
- [197] G. Pan, Z. Wu and Y. Pan. *Automatic 3D face verification from range data*. In IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, pp. 193-196, 2003.
- [198] Y. Lee and J. Shim. *Curvature based human face recognition using depth weighted Hausdorff distance*. In International Conference on Image Processing, pp. 1429-1432, 2004.
- [199] T. D. Russ, K. W. Koch and C. Q. Little. *3D facial recognition: a quantitative analysis*. In Annual Meeting of the Institute of Nuclear Materials Management, 2004.

- [200] X. Lu and A. K. Jain. *Deformation Analysis for 3D Face Matching*. In IEEE Workshop on Applications of Computer Vision, pp. 99-104, 2005.
- [201] X. Lu, D. Colbry and A. K. Jain. *Three-dimensional model based face recognition*. In *International Conference on Pattern Recognition*, pp. 362-366, 2004.
- [202] X. Lu and A. K. Jain. *Deformation modeling for robust 3D face matching*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 1377-1383, 2006.
- [203] X. Lu and A. K. Jain. *Deformation modeling for robust 3D face matching*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 8, pages 1346_1357, Aug. 2008.
- [204] F. R. Al-Osaimi, M. Bennamoun and A. S. Mian. *An expression deformation approach to non-rigid 3D face recognition*. International Journal of Computer Vision, vol. 81, no. 3, pp. 302-316, 2009.
- [205] Z. Wu, Y. Wang and G. Pan. *3D face recognition using local shape map*. In IEEE International Conference on Image Processing, volume 3, pages 2003_2006, Oct. 2004.
- [206] U. Castellani, M. Cristani, X. Lu, V. Murino and A. K. Jain. *HMM-based geometric signatures for compact 3D face representation and matching*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Biometrics, pp. 1-6, Jun. 2008.
- [207] F. Daniyal, P. Nair and A. Cavallaro. *Compact signatures for 3D face recognition under varying expressions*. In IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 302-307, Sep. 2009.
- [208] C. Maes, T. Fabry, J. Keustermans, D. Smeets, P. Suetens and D. Vandermeulen. *Feature detection on 3D face surfaces for pose normalisation and recognition*. In IEEE International Conference on Biometrics: Theory Applications and Systems, pp. 1-6, 2010.
- [209] S. Berretti, A. d.Bimbo, and P. Pala, *Sparse Matching of Salient Facial Curves for Recognition of 3-D Faces With Missing Parts*, IEEE Transactions on Information Forensics and Security, vol. 8, no. 2, 2013.
- [210] C. Samir, A. Srivastava, M. Daoudi, *Three-Dimensional Face Recognition Using Shapes of Facial Curves*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 11, pp. 1858-1863, 2006.
- [211] A. Klassen, A. Srivastava, W. Mio, S. H. Joshi, *Analysis of Planar Shapes Using Geodesic Paths on Shape Spaces*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 3, pp. 372-383, 2004.
- [212] L. Zhang, A. Razdan, G. E. Farin, J. Femiani, M. Bae and C. Lockwood. *3D face authentication and recognition based on bilateral symmetry analysis*. The Visual Computer, vol. 22, no. 1, pp. 43-55, 2006.
- [213] S. Jahanbin, H. Choi, Y. Liu and A. C. Bovik. *Three dimensional face recognition using iso-geodesic and iso-depth curves*. In IEEE International Conference on Biometrics: Theory, Applications and Systems, pp. 1-6, Oct. 2008.
- [214] H. Drira, B. Ben Amor, M. Daoudi and A. Srivastava. *Pose and expression-invariant 3D face recognition using elastic radial curves*. In British Machine Vision Conference, pp. 90.1-90.11, 2010.
- [215] N. Alyuz, B. Gokberk and L. Akarun. *Regional registration for expression resistant 3D face recognition*. IEEE Transactions on Information Forensics and Security, vol. 5, no. 3, pp. 425-440, 2010.
- [216] B. Gokberk, M. O. Irfanoglu and L. Akarun. *3D shape-based face representation and feature extraction for face recognition*. Image and Vision Computing, vol. 24, no. 8, pp. 857-869, 2006.
- [217] B. Gokberk, H. Dutagaci, A. Ulas, L. Akarun and B. Sankur. *Representation Plurality and Fusion for 3-D Face Recognition*. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 38, no. 1, pp. 155-173, 2008.
- [218] L.J. Spreeuwiers. *Fast and Accurate 3D Face Recognition Using Registration to an Intrinsic Coordinate System and Fusion of Multiple Region classifiers*. International Journal of Computer Vision, vol. 93, no. 3, pp. 389-414, 2011.

- [219] Martin A. Fischler and Robert C. Bolles. *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*. Commun. ACM, vol. 24, no. 6, pp. 381-395, 1981.
- [220] C.C. Queirolo, L. Silva, O.R.P. Bellon and M.P. Segundo. *3D Face Recognition Using Simulated Annealing and the Surface Interpenetration Measure*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 2, pp. 206-219, 2010.
- [221] S. Gupta, M. P. Sampat and Z. Wang. *Facial range image matching using the complex wavelet structural similarity metric*. In IEEE International Workshop on Applications of Computer Vision, page 4, Feb. 2007.
- [222] Y. Huang, Y. Wang and T. Tan. *Discriminating 3D faces by statistics of depth differences*. In Asian Conference on Computer Vision, vol. 2, pp. 690-699, 2007.
- [223] Y. Wang, X. Tang, J. Liu, G. Pan and R. Xiao. *3D face recognition by local shape difference boosting*. In European Conference on Computer Vision, pp. 603-616, 2008.
- [224] Y. Wang, J. Liu and X. Tang. *Robust 3D face recognition by local shape difference Boosting*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 10, pp. 1858-1870, 2010.
- [225] H. Tang, B. Yin, Y. Sun , Y. Hu, *3D face recognition using local binary patterns*, Signal Processing, vol. 93, pp. 2190-2198, 2013.
- [226] G. G. Gordon. *Face recognition from depth and curvature*. PhD thesis, Harvard University, 1992.
- [227] B. Gokberk, A. A. Salah and L. Akarun. *Rank-based decision fusion for 3D shape-based face recognition*. In International Conference on Audio- and Video-based Biometric Person Authentication, pp. 1019-1028, 2005.
- [228] M. L. Koudelka, M. W. Koch and T. D. Russ. *A prescreener for 3D face recognition using radial symmetry and the Hausdorff fraction*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Face Recognition Grand Challenge Experiments, pp. 168-168, June 2005.
- [229] Y. Lee, H. Song, U. Yang, H. Shin and K. Sohn. *Local feature based 3D face recognition*. In International Conference Audio- and Video-Based Biometric Person Authentication, pp. 909-918, 2005.
- [230] A. B. Moreno, A. Sanchez, J. Velez and J. Diaz. *Face recognition using 3D local geometrical features: PCA vs. SVM*. In IEEE International Symposium on Image and Signal Processing and Analysis, pp. 185-190, Sep. 2005.
- [231] X. Li and H. Zhang. *Adapting geometric attributes for expression-invariant 3D face recognition*. In IEEE International Conference on Shape Modeling and Applications, pp. 21-32, Jun. 2007.
- [232] P. Guan and L. Zhang. *3D face recognition based on facial structural angle and local region map*. In IEEE International Conference on Multimedia and Expo, pp. 41-44, Apr. 2008.
- [233] A. F. Abate, M. Nappi, S. Ricciardi and G. Sabatino. *Fast 3D face recognition based on normal map*. In IEEE International Conference on Image Processing, vol. 2, pp. 946-949, Sep. 2005.
- [234] A. M. Bronstein, M. M. Bronstein and R. Kimmel. *Three-dimensional face recognition*. International Journal of Computer Vision, vol. 64, no. 1, pp. 5-30, 2005.
- [235] G. Pan, S. Han, Z. Wu and Y. Wang. *3D face recognition using mapped depth images*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Face Recognition Grand Challenge Experiments, 2005.
- [236] G. Passalis, I. Kakadiaris, T. Theoharis, G. Toderici and N. Murtuza. *Evaluation of 3D face recognition in the presence of facial expressions: an annotated deformable model approach*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Face Recognition Grand Challenge Experiments, pp. 171-171, 2005.
- [237] R. S. Llonch, E. Kokiopoulou, I. Tosic and P. Frossard. *3D face recognition using sparse spherical representations*. In International Conference on Pattern Recognition, pp. 1-4, 2008.

- [238] T. C. Faltemier, K. W. Bowyer and P. J. Flynn. *A region ensemble for 3-D face recognition*. IEEE Transactions on Information Forensics and Security, vol. 3, no. 1, pp. 62-73, 2008.
- [239] N. Alyuz, B. Gookberk and L. Akarun. *Regional registration for expression resistant 3-D face recognition*. IEEE Transactions on Information Forensics and Security, vol. 5, no. 3, pp. 425-440, 2010.
- [240] C. Beumier, M. Acheroy. *Automatic 3D Face Authentication*. Image and Vision Computing, vol. 18, no. 4, pp. 315-321. 2000.
- [241] Y. Wang, C. S. Chua, and Y. K. Ho. *Facial feature detection and face recognition from 2d and 3d images*. Pattern Recognition Letters, vol. 23, no. 10, pp. 1191-1202, 2004.
- [242] C. S. Chua, F. Han, and Y. K. Ho. *3D human face recognition using point signature*. In Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, vol. 1, pp. 233-238. 2000.
- [243] A.M. Bronstein, M.M. Bronstein, E. Gordon, R. Kimmel, *Fusion of 2D and 3D in Three-Dimensional Face Recognition*. in IEEE International Conference on Image Processing, pp. 87-90, 2004.
- [244] X. Lu, D. Colbry, and A.K. Jain. *Matching 2.5 D scans for face recognition*. pp. 30-36. Springer, 2004.
- [245] D. M. Weinstein. *The Analytic 3-D Transform for the Least-Squared Fit of Three Pairs of Corresponding Points*. School of Computing Technical Report, pp. 98-005.
- [246] M. Hüsken, M. Brauckmann, S. Gehlen, C. V. der Malsburg, *Strategies and Benefits of Fusion of 2D and 3D Face Recognition*, in CVPR'05 : the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, pp. 174. 2005.
- [247] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, G. Medioni, *Performance of Geometrix ActiveIDTM 3D Face Recognition Engine on the FRGC Data*, in CVPR 05 : the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, pp. 154. 2005.
- [248] G. Mediouni, R. Waupotitsch, *Face recognition and modeling in 3D*, in IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp. 232-233. 2003.
- [249] Y. Wang and C.S. Chua. *Face recognition from 2D and 3D images using 3D Gabor filters*. Image and Vision Computing, vol. 23, no. 11, pp. 1018-1028, 2005.
- [250] D.G. Sim, O.K. Kwon, and R.H. Park. *Object matching algorithms using robust Hausdorff distance measures*. IEEE Transactions on Image Processing, vol. 8, no. 3, pp. 425-429, 1999.
- [251] J. Cook, C. McCool, V. Chandran, and S. Sridharan. *Combined 2D/3D Face Recognition Using Log-Gabor Templates*. In Proceedings of the IEEE International Conference on Video and Signal Based Surveillance, pp. 83, 2006.
- [252] I. Mpiperis, S. Malassiotis, and M. G. Strintzis. *3-D Face Recognition With the Geodesic Polar Representation*. IEEE Transactions on Information Forensics and Security, vol. 2, no. 3, pp. 537-547, 2007.
- [253] C. Xu, S. Li, T. Tan, and L. Quan. *Automatic 3D face recognition from depth and intensity Gabor features*. Pattern Recognition, vol. 42, no. 9, pp. 1895-1905, 2009.
- [254] Y. Freund and R. Schapire. *A decision-theoretic generalization of on-line learning and an application to boosting*. In Computational Learning Theory, pp. 23-37. Springer, 1995.
- [255] F. R. Al-Osaimi, M. Bennamoun, and A. Mian. *Spatially Optimized Data-Level Fusion of Texture and Shape for Face Recognition*. IEEE Transactions on Image Processing, vol. 21, no. 2, 2012.
- [256] X. Wang, Q. Ruan, G. An, Y. Jin. *2D+3D Face Recognition Using Dual-tree Wavelet Transform*. TENCON IEEE Region 10 Conference, pp. 1-4, 2013.
- [257] A. Mellakh. *Reconnaissance des visages en conditions dégradées*. THÈSE pour l'obtention du grade de Docteur de l'Institut National des Télécommunications Spécialité : Traitement d'image et de signal, 07 Avril 2009.

- [258] C. G. Rafael and E. W. Richard. *Digital Image Processing* (2nd Edition). Prentice Hall, January 2002.
- [259] W. Hizem. *Capteur Intelligent pour la Reconnaissance de Visage*. Thèse Doctorat délivré conjointement par L'Institut National des Télécommunications et l'Université Pierre et Marie Curie - Paris 6 Spécialité : Electronique /Informatique, 2009.
- [260] L. D. Ruderman. *The statistics of natural images. Network : Computation in Neural Systems*, pp. 517-548, 1994.
- [261] P.S. Penev and J.J. Atick. *Local feature analysis : A general statistical theory for object representation*. *Network : Computation in Neural Systems*, vol. 7, no. 3, pp. 477-500, 1996.
- [262] D. Saigaa, *Contribution à l'authentification d'individus par reconnaissance de visages*, thèse de Doctorat d'état en automatique, Université Mohamed khider Biskra Algérie, Novembre 2006.
- [263] C. Liu and H. Wechsler, *Robust coding schemes for indexing and retrieval from large face databases*, *IEEE Trans. on Image Processing*, vol. 9, no. 1, pp. 132-137, 2000.
- [264] Hua Yu and Jie Yang. *A direct lda algorithm for high-dimensional data with application to face recognition*. *Pattern Recognition*, vol. 34, no. 10, pp. 2067-2070, 2001.
- [265] W. Hwang, H. Wang, H. Kim, S. C. Kee, and J. Kim. *Face Recognition System Using Multiple Face Model of Hybrid Fourier Feature under Uncontrolled Illumination Variation*. *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 1152- 1165, 2011.
- [266] K. Messer, J. Matas, J. Kittler, J. Luetttin, and G. Maitre. *XM2VTSDB: The Extended M2VTS Database*. In *Proceedings, International Conference on Audio- and Video-Based Person Authentication*, pp. 72-77, 1999.
- [267] A. V. Oppenheim and J. S. Lim. *The importance of phase in signals*. *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529-541, 1981.
- [268] C. Liu. *Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance*. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 725-737, May 2006.
- [269] L.L. Shen and L. Bai. *Gabor feature based face recognition using kernel methods*. In *AFGR04*, pp.170-175, 2004.
- [270] M. Zhou and H. Wei. *Face verification using gabor wavelets and adaboost*. In *IEEE Computer Society Proceedings of the 18th International Conference on Pattern Recognition*, Washington, DC, USA, pp. 404-407, 2006.
- [271] D. P. Delacrétaz, G. Chollet, and B. Dorizzi. *Guide to Biometric Reference Systems and Performance Evaluation*. Springer, 2009.
- [272] A. K. Jain and A. Ross. *Multibiometric systems. Communications of the ACM*, special issue on multimodal interfaces, vol. 47, no. 1, pp. 34-40, 2004.
- [273] Y. Chen, S. Dass, and A. Jain. *Fingerprint Quality Indices for Predicting Authentication Performance*. In : *Proceedings of Fifth International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, New York, NY, USA, pp. 160- 170, July 2005.
- [274] NIST report to the United States Congress. *Summary of NIST Standards for Biometric Accuracy, Tamper Resistance, and Interoperability*. November 2002. Available at ftp://sequoyah.nist.gov/pub/nist_internal_reports/NISTAPP_Nov02.pdf.
- [275] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. *Impact of Artificial "Gummy" Fingers on Fingerprint Systems*. In : *Proceedings of SPIE : Optical Security and Counterfeit Deterrence Techniques IV*, pp.275-289, January 2002.
- [276] L. ALLANO. *La Biométrie multimodale : stratégies de fusion de scores et mesures de dépendance appliquées aux bases de personnes virtuelles*. Thèse présentée pour l'obtention du grade de Docteur de l'Institut National des Télécommunications, 12 janvier 2009.
- [277] M. Haindl, J. Kittler, F. Roli. (eds.): *MCS 2007*. LNCS, vol. 4472. Springer, 2007.
- [278] Z. Yu, M. Y. Nam, S. Sedai, and P. K. Rhee. *Evolutionary Fusion of a Multi-Classifer System for Efficient Face Recognition*. *International Journal of Control, Automation, and Systems*, vol. 7, no. 1, pp. 33-40. 2009.

- [279] X. Zhang, D. Liu, and J. Chen. *An Illumination Independent Face Verification Based on Gabor Wavelet and Supported Vector Machine*. Communications in Computer and Information Science, vol. 15, pp. 153-160, 2008.
- [280] Z. S. Zhao, J. Z. Wang, Q. J. Tian, J. M. Zhou. *Classifier Fusion of Improved Laplacianfaces*. IEEE Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, pp. 11-14 July 2010.
- [281] Y. Wang, T. Tan, and A. Jain. *Combining face and iris biometrics for identity verification*. In: Proceedings of Fourth International Conference on Audio- and Video-Based Authentication (AVBPA), pp. 805-813, June 2003.
- [282] P. Verlinde, P. Druyts, G. Cholet, and M. Acheroy. *Applying Bayes based classifiers for decision fusion in a multi-modal identity verification system*. In: Proceedings of International Symposium on Pattern Recognition, February 1999.
- [283] V. Chatzis, A. Bors, and I. Pitas. *Multimodal decision-level fusion for person authentication*. IEEE Transactions on Systems, Man and Cybernetics, Part A : Systems and Humans, vol. 29, no. 6, pp. 674-681, November 1999.
- [284] C. Sanderson and K. Paliwal. *Information fusion and person verification using speech and face information*. Tech. Rep. IDIAP-RR 02-33, IDAIP, September 2002.
- [285] A. Ross and A. Jain. *Information fusion in biometrics*. Pattern Recognition Letters, vol. 24, no. 13, pp. 2115-2125, 2003.
- [286] J. Kittler, M. Hatef, P. R. Duin, J. Matas. *On combining classifiers*. IEEE Trans. Pattern Analysis and Mach. Intell. vol. 20, no. 3, pp. 226-239, 1998.
- [287] K. C. Kwak and W. Pedrycz. *Face recognition: A study in information fusion using fuzzy integral*. Patter Recognition Letter, vol. 26, pp. 719-733, 2005.
- [288] C. Chia, N. Sherkat, and L. Nolle. *Confidence Partition and Hybrid Fusion in Multimodal Biometric Verification System*. Lecture Notes in Computer Science, vol. 5707, pp. 212-219, 2009.
- [289] Y. Li, J. Yin, E. Zhu, C. Hu, and H. Chen. *Studies of Fingerprint Template Selection and Update*. Workshops and Symposia, CCIS 28, pp.150-163, 2009.
- [290] A. Shukla et al. *Multimodal Biometric Systems*. Chapter 18: Towards Hybrid and Adaptive Computing, Springer, pp.401-418, 2010.
- [291] R. Brunelli and D. Falavigna. *Person identification using multiple cues*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol17, pp.955-966, 1995.
- [292] S. Tulyakov, S. Jaeger, and V. Govindaraju, and D. Doermann. *Review of Classifier Combination Methods*. Machine Learning in Document Analysis and Recognition, vol. 90, pp.361-386, 2008.
- [293] A. A. Ross, K. k Nandakumar, and A. K. Jain. *Handbook of Multibiometrics*. International Series on Biometrics, vol. 6, 2006.
- [294] R. Snelick and M. Indovina. *Large-Scale Evaluation of Multimodal Biometric Authentication Using State-of-the-Art Systems*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pp. 450-455, 2005.
- [295] R. Cappelli, D. Maio, and D. Maltoni. *Combining Fingerprint Classifiers*. In: Proceedings of the First International Workshop on Multiple Classifier Systems, pp.351-361, Springer, 2000.
- [296] F. WANG and J. HAN, *Multimodal biometric authentication based on score level fusion using support vector machine*, Opto-Electron. Vol. 17, no. 1, 2009.
- [297] L. A. Zadeh. *Fuzzy sets*. Information Control. vol. 8, pp. 338-353,1965.
- [298] M. Sugeno. *Fuzzy measures and fuzzy integrals A survey*. In Fuzzy Automata and Decision Processes, pp. 89-102, 1977.
- [299] Ke. C. Kwak and W. Pedrycz. *Face Recognition Using Fuzzy Integral and Wavelet Decomposition Method*. IEEE Transactions on Systems , Part B: Cybernetics, vol. 34, no. 4, 2004
- [300] A. M.P. Canuto and M. C.C. Abreu. *Using Fuzzy*. Neural and Fuzzy-Neural Combination Methods in Ensembles with Different Levels of Diversity, Artificial Neural Networks, vol. 4668, 2007, pp. 349-359, 2007.

- [301] T. Murofushi and M. Sugeno, *An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure*. Fuzzy Sets Syst. vol. 29, pp. 201-227, 1988.
- [302] P. Verlinde. *A contribution to multi-modal identity verification using decision fusion*. PhD thesis, Ecole Nationale des Télécommunications France, 1999.
- [303] B. Duc, E. S. Bigün, J. Bigün, G. Maître, and S. Fischer. *Fusion of audio and video information for multi modal person authentication*. Pattern Recognition Letters, vol. 18, no. 9, pp. 835-843, 1997.
- [304] E. S. Bigün, Josef Bigün, Benoît Duc, and S. Fischer. *Expert conciliation for multi modal person authentication systems by bayesian statistics*. Proceedings of the First International Conference on Audioand Video-Based Biometric Person Authentication, pp.291-300, 1997.
- [305] S. Pigeon, *Authentification Multimodale D'Identité*. Thèse présentée en vue de l'obtention du grade de Docteur en Sciences Appliquées. Laboratoire de Télécommunications et Télédétection Place du Levant, Belgique, Février 1999.
- [306] V. N. Vapnik. *The nature of statistical learning theory*. Book: The nature of statistical learning theory, 1995.
- [307] D. A. Ramli, S. A. Samad, and A. Hussain. *Score Information Decision Fusion Using Support Vector Machine for a Correlation Filter Based Speaker Authentication System*. Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems CISIS'08, vol. 53, pp. 235-242, 2009.
- [308] P. Ejarque, J. Hernado, D. Hernando, and D. Gómez. *Eigenfeatures and Supervectors in Feature and Score Fusion for SVM Face and Speaker Verification*. Biometric ID Management and Multimodal Communication, vol. 5707, pp. 81-88, 2009.
- [309] D. R. Kisku, P. Gupta, and J. K. Sing. *Fusion of Multiple Matchers Using SVM for Offline Signature Identification*. Security Technology, vol. 58, pp. 201-208, 2009.
- [310] M. Farrús, P. Ejarque, A. Temko, and J. Hernando. *Histogram Equalization in SVM Multimodal Person Verification*. Advances in Biometrics, vol. 4642, pp. 819-827, 2007.
- [311] R. Benmokhtar and B. Huet. *Classifier Fusion: Combination Methods For Semantic Indexing in Video Content*. Artificial Neural Networks, vol. 4132, pp. 65-74, 2006.
- [312] M. ARIF. *Fusion de Données : Ultime Etape de Reconnaissance de Formes*. Applications à l'Identification et à l'Authentification. Thèse pour obtenir le grade de Docteur de l'université de Tours, 25 novembre 2005.
- [313] P. J. Besl and N. D. McKay. *A method for registration of 3-d shapes*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 14, no. 2, 239-256, 1992.
- [314] Y. Chen and G. Medioni. *Object modelling by registration of multiple range images*. Image Vision Comput., vol. 10, no. 3, pp. 145-155, 1992.
- [315] M. Belahcene, A. Ouamane, M. Boumehrez and A. Benakcha. *Comparaison Des Methodes De Reduction D'espace Et L'application Des SVMs Pour La Classification Dans L'authentification De Visages*. N°13, pp.116-126, 2012.
- [316] A. Ouamane, M. Belahcene, A. Benakcha, M. Boumehrez, A.Taleb Ahmed. *Identification of Faces by Multimodal Information Fusion of Depth and Color*. Sensors & Transducers Journal, vol. 140, no. 5, pp. 74-87, 2012.
- [317] D. Harwood, T. Ojala, M. Pietikäinen, S. Kelman, and S. Davis. *Texture Classification by Center-Symmetric Auto-correlation, Using Kullback Discrimination of Distributions*. Technical Report, CAR-TR-678, Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park, Maryland, 1993.
- [318] A. Ouamane, M. Belahcene, A. Benakcha, M. Boumehrez, A. T. Ahmed. *The Classification of Scores from Multi-classifiers for Face Verification*. Sensors and Transducers Journal, vol. 145, no. 10, pp. 116-118, 2012.
- [319] D. Lowe. *Distinctive Image Features from Scale-Invariant Key Points*. Intl J. Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

- [320] D. Huang, M. Ardabilian, Y. Wang, L. Chen. 3-D Face Recognition Using eLBP-Based Facial Description and Local Feature Hybrid Matching. *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1551-1565, 2012.
- [321] D. Lowe. Demo Software: SIFT Keypoint Detector. <http://www.cs.ubc.edu.ca/lowe/>, 2006.
- [322] *Identification of Faces by Multimodal Information*
- [323] C. Xu, Y. Wang, T. Tan and L. Quan. 3D Face Recognition Based on G-H Shape Variation. *Lecture Notes in Computer Science*, vol. 3338, pp. 233-243, 2005.
- [324] A. Ouamane & M. Belahcene. *New Multiple Classifier Systems For Face Authentication*. *Courrier du Savoir*, N°18, pp.53-62, Mars 2014.
- [325] J. Flusser et T. Suk. *Degraded Image Analysis: An Invariant Approach*. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 590-603, 1998.
- [326] C. Fiche. *Repousser les limites de l'identification faciale en contexte de vidéosurveillance*. Docteur De L'université De Grenoble Spécialité : Signal - Images - Parole - Télécoms (SIPT), le 31/01/2012.
- [327] M. Banham, A. Katsaggelos. *Digital image restoration*. *IEEE Signal Process. Mag.* vol. 14, no. 2, no. 24-41, 1997.
- [328] V. Ojansivu, J. Heikkilä. *A method for blur and similarity transform invariant object recognition*. In: *Proceedings of International Conference on Image Analysis and Processing*, pp. 583-588, 2007.
- [329] S. Brahmam, L. C. Jain Loris Nanni, A. Lumini. *Local Binary Patterns: New Variants and Applications*. Springer: *Studies in Computational Intelligence*, vol. 506, 2014.
- [330] B. Zhang, S. Shan, X. Chen, and W. Gao. *Histogram of gabor phase patterns (hgpp): A novel object representation approach for face recognition*. *TIP*, vol. 16, no. 1, 2007.
- [331] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkilä. *Recognition of blurred faces using local phase quantization*. In *ICPR*, pp. 1-4, 2008.
- [332] A. Hyvärinen et al. *Natural Image Statistics*. Springer, 2009.
- [333] A. Hyvärinen and E. Oja. *Independent component analysis: algorithms and applications*. *Neural Networks*, 2000.
- [334] <http://bosphorus.ee.boun.edu.tr/>.
- [335] X. Wang, Q. Ruan, Y. Ming. *3D Face Recognition Using Corresponding Point Direction Measure and Depth Local Features*. *IEEE Int. Conf. on Signal Processing*, pp. 86-89, 2010.
- [336] Y. A. Li, Y. Jun, S. G. Zhang, H. L. Xu. *An Efficient 3D Face Recognition Method Using Geometric Features*. *IEEE Int. Conf. on Intelligent Systems and Applications*, pp. 1-4, 2010.
- [337] Y. Jin , Y. Wang, Q. Ruan, X. Wang. *A New Scheme for 3D Face Recognition Based on 2D Gabor Wavelet Transform Plus LBP*. *IEEE Int. Conf. on Computer Science & Education*, pp. 860-865, 2011.
- [338] S. Berretti, N. Werghi, A. d. Bimbo, P. Pala. *Matching 3D face scans using interest points and local histogram descriptors*. *Computers & Graphics*, vol. 37, no. 5, pp. 509-525, 2013.
- [339] D. Smeets, J. Keustermans, D. Vandermeulen, and P. Suetens. *meshSIFT: Local surface features for 3D face recognition under expression variations and partial data*. *Computer Vision and Image Understanding*, vol. 117, pp.158-169, 2013.
- [340] X. Zhao, E. Dellandréa , J. Zou , and L. Chen. *A unified probabilistic framework for automatic 3D facial expression analysis based on a Bayesian belief inference and statistical feature models*. *Image and Vision Computing*, vol. 31, pp. 231-245, 2013
- [341] Y. Ming. *Rigid-area orthogonal spectral regression for efficient 3D face recognition*. *Neurocomputing*, vol. 129, no. 10, pp. 445-457, 2014.
- [342] M. Pietikinen, A. Hadid, G. Zhao, T. Ahonen. *Computer Vision Using Local Binary Patterns*. *Computational Imaging and Vision*, vol. 40, 2011.
- [343] A. Ouamane, M. Bengherabi, A. Guessoum, A. Hadid and M. Cheriet. *Efficient face verification system based on multiscale multidesciptor local binary features and exponentiel discriminant analysis*. *ICIP*, 2014.

- [344] C. Ho Chan, M. A. Tahir, J. Kittler, and M. Pietikäinen. *Multiscale Local Phase Quantization for Robust Component-Based Face Recognition Using Kernel Fusion of Multiple Descriptors*. IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 5, pp. 1164-1177, 2013.
- [345] Y. Rodriguez and S. Marcel, *Face authentication using adapted local binary pattern histograms*. ECCV, pp. 321-332, 2006.
- [346] G. Heusch, Y. Rodriguez, and S. Marcel. Local binary patterns as an image preprocessing for face authentication. 7th International Conference on Automatic Face and Gesture Recognition, pp. 9-14, 2006.
- [347] K. Messer, J. Kittler, J. Short, G. Heusch, F. Cardinaux, S. Marcel, Y. Rodriguez, S. Shan, Y. Su, W. Gao, and X. Chen. *Performance characterisation of face recognition algorithms and their sensitivity to severe illumination changes*. ICB, pp. 1-11, 2006.
- [348] <http://vis-www.cs.umass.edu/lfw/>
- [349] L. Wolf, T. Hassner, and Y. Taigman. *Effective Face Recognition by Combining Multiple Descriptors and Learned Background Statistics*. IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 10, pp. 1978-1990, 2011.
- [350] Y. Taigman and L. Wolf. *Leveraging Billions of Faces to Overcome Performance Barriers in Unconstrained Face Recognition*. CoRR, pp. 1108-1122, 2011.
- [351] X. Tan, B. Triggs. *Enhanced local texture feature sets for face recognition under difficult lighting conditions*. IEEE Trans. Image Process, vol. 19, 2010.
- [352] A. O. Hatch, S. Kajarekar and A. Stolcke. *Within-class covariance normalization for SVM-based speaker recognition in Proc. ICSLP*, pp. 1471-1474, 2006.
- [353] M. Belahcene, A. Ouamane, A. T. Ahmed. *Fusion by Combination of Scores Multi-Biometric Systems*. in Proc. IEEE Conf. EUVIP, pp. 252-257, July. 2011.
- [354] X. Li, F. Da. *Efficient 3D face recognition handling facial expression and hair occlusion*. Image and Vision Computing, vol. 30, no. 9, pp. 668- 679, 2012.
- [355] C. Dorai, A.K. Jain. *COSMOS-A representation scheme for 3D free-form objects*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 10, pp. 1115-1130, 1997.
- [356] F.B. Haar, R.C. Veltkamp. *A 3D face matching framework for facial curves*. Graph. Model., vol. 71, no. 2, pp. 77-91, 2009.
- [357] N. L. Shchegoleva and G. A. Kukharev. *Application of Two Dimensional Principal Component Analysis for Recognition of Face Images*. Pattern Recognition and Image Analysis, vol. 20, no. 4, pp. 513-527, 2010.
- [358] T. Mäenpää, T. Ojala, M. Pietikäinen, and S. Maricor. *Robust texture classification by subsets of local binary patterns*. In Pattern Recognition, International Conference on, vol 3, pp. 3947, 2000.
- [359] R. Sidje and W. Stewart. *A numerical study of large sparse matrix exponentials arising in Markov chains*. Comput. Statist. Data Anal., vol. 29, no. 3, pp. 345-368, 1999.
- [360] S. J. Wang, S. Yan, J. Yang, C. G. Zhou, and X. Fu. *A General Exponential Framework for Dimensionality Reduction*. IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, 2014.
- [361] G. H. Golub and C. F. V. Loan. *Matrix Computation*. Baltimore, MD: The Johns Hopkins Univ. Press, 1983.
- [362] K. Z. Mao. *Fast orthogonal forward selection algorithm for feature subset selection*. IEEE Trans. Neural Netw., vol. 13, no. 5, pp. 1218-1224, Sep. 2002.
- [363] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet. *Front-end factor analysis for speaker verification*. IEEE Trans. on Audio, Speech, and Language Processing, vol. 19, no. 4, pp. 788-798, 2011.

- [364] A. Mian, M. Bennamoun, R. Owens. *An efficient multimodal 2D 3D hybrid approach to automatic face recognition*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 11, pp.1927-1943, 2007.
- [365] P. Shih and C. Liu. *Wavelet Features for 3D Face Recognition Cross Disciplinary Biometric Systems*. pp. 93-116, 2012.
- [366] A. S. Mian, M. Bennamoun, R. Owens, *Keypoint detection and local feature matching for textured 3D face recognition*, Computer Vision, vol. 79, no. 1, pp. 1-12, 2008.
- [367] C. Xu, S. Li, T. Tan and L. Quan. *Keypoint Automatic 3D face recognition from depth and intensity Gabor features*. Pattern Recognition , vol. 42, no. 9, pp. 1895-1905, 2009.
- [368] G. Zhang, Y. Wang. *Robust 3D face recognition based on resolution invariant features*. Pattern Recognition Letters, vol. 32, no. 7, pp. 1009-1019, 2011.
- [369] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer. *Overview of the face recognition grand challenge*. in: Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 947-954, 2006.
- [370] A. Mian ,Bronstein, M. M. Bronstein, R. Kimmel. *Expression-invariant representations of faces*. IEEE Trans. Image Process., vol. 16, no.1, pp.188-197, 2007.
- [371] A. Ouamane, M. Belahcne, S. Bourennane. *Multimodal 3d and 2d Face Authentication Approach Using Extended LBP and Statistic Local Features Proposed*. in Proc. IEEE Conf. EUVIP, June. 2013, pp. 130-135.
- [372] H. Lee, A. Battle, R. Raina, and A. Y. Ng. *Efficient sparse coding algorithms*. In Advances in Neural Information Processing Systems 20, NIPS, 2006.
- [373] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. *Online dictionary learning for sparse coding*. In Proceedings of the International Conference on Machine Learning, ICML, 2009.
- [374] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. *Supervised dictionary learning*. In Advances in Neural Information Processing Systems 23, NIPS, 2009.
- [375] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, and D. Cai. *Graph regularized sparse coding for image representation*. IEEE Transactions on Image Processing, 20(5), 2011.
- [376] H. Nguyen and L. Bai. *Cosine Similarity Metric Learning for Face Verification*. Proc. Asian Conf. Computer Vision, pp. 709-720, 2011.
- [377] K. Weinberger, J. Blitzer, and L. Saul. *Distance Metric Learning for Large Margin Nearest Neighbor Classification*. Proc. Conf. Advances in Neural Information Processing Systems, 2005.
- [378] J. Davis, B. Kulis, P. Jain, S. Sra, and I. Dhillon. *Information-Theoretic Metric Learning*. Proc. Int'l Conf. Machine Learning, pp. 209-216, 2007.
- [379] Jiwen Lu, Xiuzhuang Zhou, Yap-Pen Tan, Yuanyuan Shang, and Jie Zhou. *Neighborhood Repulsed Metric Learning for Kinship Verification*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 36, No. 2, February 2014.