

Université Mohamed Khider Biskra
Faculté des Sciences et de la Technologie
Département : Génie Electrique
Ref :.....



جامعة محمد خيضر بسكرة
كلية العلوم والتكنولوجيا
قسم: الهندسة الكهربائية
المرجع:.....

Thèse présentée en vue de l'obtention
du diplôme de
Doctorat LMD en Génie Electrique

Filière : Electronique

Option : Signaux et Communication

**Analyse d'images d'expressions faciales et orientation
de la tête basée sur la profondeur**

Présentée par :

CHOUCHANE Ammar

Soutenue publiquement le : 04/06/2016

Devant le jury composé de :

Dr. DEBILOU Abderrazak	Professeur	Président	Université de Biskra
Dr. BELAHCENE Mébarka	M. C. A	Rapporteur	Université de Biskra
Dr. ABDOU Latifa	M. C. A	Examinatrice	Université de Biskra
Dr. BOUTARFA Abdelhalim	Professeur	Examineur	Université de Batna
Dr. BOURENNANE Salah	Professeur	Examineur	Ecole Centrale de Marseille

Dédicaces

A tout respect et amour je dédie ce travail à mes très chers parents, à ma chère épouse, ma famille et tous mes amis

Ammar

Remerciements

Tout d'abord, je tiens à exprimer mes remerciements à ma directrice de thèse Mme **BELAHCENE** Mébarka, Maitre de Conférences A à l'Université Mohamed **KHIDER** Biskra, de m'avoir soutenu et fait confiance durant ces années avec une grande patience. Avec son expérience dans la recherche et l'enseignement, avec ses conseils, j'ai pu découvrir le monde de la recherche scientifique dans le domaine du traitement d'image et des techniques de la biométrie faciale. Il est effectivement un grand plaisir pour moi de travailler sous la supervision de Mme **BELAHCENE** Mébarka.

Je tiens à remercier chaleureusement Mr **OUAMANE** Abdelmalik, enseignant chercheur au laboratoire d'identification, de commande, contrôle et communication Biskra, pour sa contribution, son aide et ses encouragements qu'il m'a prodigués durant la période de la thèse.

Merci, bien évidemment à Mr **BOURENNANE** Salah Professeur à l'Ecole Centrale de Marseille en France et Responsable du GSM, Groupe Signaux Multidimensionnels pour son soutien, ses conseils fructueux ainsi que pour l'accueil dans son laboratoire.

Je remercie aussi Mr **MAROT** Julien, Maitre de Conférences à l'Ecole Centrale de Marseille pour son aide durant mon stage au sein du groupe GSM nos discussions ont permis d'explorer efficacement certaines pistes.

Je tiens à remercier profondément les membres du jury :

1. Mr **DEBILOU** Abderrazak, Professeur à l'Université de Biskra, trouve ici l'expression de mes remerciements les plus sincères pour avoir accepté de présider cette thèse.
2. Mr **BOUTARFA** Abdelhalim, Professeur à l'Université El Haj **LAKHDER** Batna pour l'intérêt porté à notre travail et sa participation au jury en tant qu'examinateur.
3. Mme **ABDOU** Latifa Maitre de Conférences A à l'Université Mohamed **KHIDER** Biskra, d'avoir accepté d'être examinatrice.

Enfin, je remercie mes amis et ma famille pour leur patience, leurs encouragements et leur compréhension concernant mon absence, surtout en fin de thèse.

Publications & Communications associées à la thèse

Les contributions présentées dans ce manuscrit de thèses ont été publiés dans les articles suivants :

I. Publications internationales

1. **A. Chouchane**, M. Belahcene, and S. Bourennane, "3D and 2D face recognition using integral projection curves based depth and intensity images," International Journal of Intelligent Systems Technologies and Applications, vol. 14, pp. 50-69, **2015**.
2. **A. Chouchane**, M. Belahcene, A. Ouamane and S. Bourennane, "Evaluation of Histograms Local Features and Dimensionality Reduction for 3D Face Verification," (**In Press**), Journal of Information Processing Systems. **2016**, Doi >10.3745/JIPS.02.0037.

II. Publication nationale

3. **A. Chouchane**, M. Belahcene, "Automatic Face recognition using Integral Projection Curves and patch of interest regions," **2015** (*Submitted*).

III. Communications internationales

4. M. Belahcene, **A. Chouchane**, and H. Ouamane, "3D face recognition in presence of expressions by fusion regions of interest," in Signal Processing and Communications Applications Conference (SIU), **2014** 22nd. IEEE, Turquie 2014, pp. 2269-2274.
5. **A. Chouchane**, M. Belahcene, A. Ouamane, and S. Bourennane, "Multimodal face recognition based on histograms of three local descriptors using score level fusion," in Visual Information Processing (EUVIP), 2014 5th European Workshop on. IEEE, Paris France **2014**, pp. 1-6.
6. **A. Chouchane**, M. Belahcene, A. Ouamane, and S. Bourennane, "3D face recognition based on histograms of local descriptors," in Image Processing Theory, Tools and Applications (IPTA), 2014 4th International Conference on. IEEE, Paris France **2014**, pp. 1-5.
7. M. Belahcene, **A. Chouchane**, and N. Mokhtari, "2D and 3D face recognition based on IPC detection and patch of interest regions," in Connected Vehicles and Expo (ICCVE), 2014 International Conference on. IEEE, Vienne Autriche **2014**, pp. 627-628.

8. M. Belahcene, **A. Chouchane**, M. Amin Benatia, and M. Halitim, "3D and 2D face recognition based on image segmentation," in Computational Intelligence for Multimedia Understanding (IWCIM), Paris France 2014 International Workshop on. IEEE, **2014**, pp. 1-5.
9. **A. Chouchane**, M. Belahcene, S. Bourenane, " 3D Face Detection using IPC Based Depth and Intensity images", in Medical Measurements and Application (MeMeA), International Symposium on IEEE, Italie **2015**.

IV. Communication nationale

10. M. Belahcene, **A. Chouchane**," Reconnaissance de visage basée sur l'analyse et les statistiques ", WIMAM'14, **2014** Guelma Algérie.

Résumé

Au cours des deux dernières décennies, le système de la reconnaissance automatique de visage est devenu l'une des tâches les plus intéressantes dans le domaine de la vision par ordinateur. Ces systèmes ont attirés beaucoup d'intérêt en raison de leurs diverses applications dans la vie humaine. La reconnaissance de visage est une approche biométrique basée sur des méthodes automatiques pour vérifier l'identité des personnes à base de modalités. Cette thèse est consacrée à la vérification automatique de visage basée sur les images de profondeurs 3D. Malheureusement, les conditions non contrôlées telles que les variations d'illuminations, les expressions faciales et les variations de poses affectent négativement les performances de ces systèmes surtout les systèmes basés sur les images 2D. L'obtention d'une performance élevée dans le monde réel reste donc un problème ouvert. Récemment, les travaux de recherche sont tournés vers la reconnaissance de visage utilisant l'information faciale 3D qui représente des caractéristiques plus discriminantes. Dans un premier temps, nous menons une étude approfondie des techniques récentes de la reconnaissance faciale 3D ainsi qu'algorithmes qui traitent les difficultés mentionnées. Ensuite, une nouvelle méthode de détection de visage est proposée : la segmentation de la zone du visage avec les courbes de projection intégrale IPC (Integral Projection Curves) suivi par la localisation de bout du nez. Les résultats obtenus sur les bases de données CASIA3DV4 et GavabDB et leurs comparaisons avec les travaux de l'état de l'art confirment tout l'intérêt de la bonne précision de la détection pour un système de reconnaissance efficace. Enfin, nous proposons un nouveau cadre pour la vérification 3D du visage en utilisant la réduction de la dimensionnalité basée sur les histogrammes des descripteurs locaux multi-blocs. À cet effet, cinq descripteurs locaux sont évalués, LBP (Local Binary Patterns), TPLBP (Three-Patch Local Binary Patterns), FPLBP (Four-Patch Local Binary Patterns), BSIF (Binarized Statistical Image Features) et LPQ (Local Phase Quantization). Des expériences sur les combinaisons de ces descripteurs au niveau des caractéristiques à l'aide d'une concaténation des histogrammes sont effectuées. La performance de l'approche proposée est évaluée avec différents algorithmes de réduction de la dimension : Principal Component Analysis (PCA), Orthogonal Locality Preserving Projection (OLPP) et PCA (Principal Component Analyses) + EFM (Enhanced Fisher linear discriminant Model). La multi-classe machine à vecteurs de support (SVM) est utilisée comme un classificateur pour effectuer la vérification entre les imposteurs et les clients.

Nous proposons aussi une nouvelle stratégie de vérification de visage 3D robuste et efficace aux grandes variations des poses en utilisant, l'apprentissage par les angles d'Euler (poses : $\pm 60^\circ$ selon l'axe Y) et la modélisation tensorielle basée sur la méthode proposée HOSVD (Higher Order Singular Value Decomposition) + PCA+EFM en utilisant l'algèbre multilinéaire.

Mots-Clés : Vérification de visage, Image de Profondeur, Segmentation par IPC, Descripteurs locaux multi-blocs, Réduction de dimension, Représentation tensorielle.

Abstract

Over the past two decades, the automatic face recognition has been become one of the most interesting tasks in the field of computer vision that has been attracted a lot of interest due to their various applications. These systems have attracted considerable interest due to their various applications in human life. This thesis is devoted to automatic face verification based on 3D depth images. The uncontrolled conditions such as illumination variations, facial expressions, and the head pose variations adversely affect the performance these systems. Therefore, obtaining a high recognition performance in real world is an open problem. Recently, the research turned to face recognition using 3D facial information, which represents the most discriminating characteristics. A new face detection method has been proposed: IPC (Integral Projection Curves) based facial area segmentation. The results obtained on CASIA3DV4 databases and GavabDB and their comparison with the work of the state of the art confirm the interest of good detection accuracy for recognition. Then, we propose a new framework for 3D face verification using the dimensionality reduction based on the histograms of multi-block local descriptors to represent the facial images distinctively. For this purpose, five different local descriptors are evaluated, LBP (Local Binary Patterns), TPLBP (Three-Patch Local Binary Patterns), FPLBP (Four-Patch Local Binary Patterns), BSIF (Binarized Statistical Image Features) and LPQ (Local Phase Quantization). Furthermore, experiments on the combinations of the five local descriptors at feature level using simply histograms concatenation are provided. The performance of the proposed approach is evaluated with different dimensionality reduction algorithms: Principal Component Analysis (PCA), Orthogonal Locality Preserving Projection (OLPP) and la combination PCA+EFM (Enhanced Fisher linear discriminant Model). Multi-class Support Vector Machine (SVM) is used as a classifier to carry out the verification between imposters and customers.

At the end of this thesis, we propose a new strategy of 3D face verification robust and efficient to large variations of the poses using, learning by Euler angles (poses: $\pm 60^\circ$ along the Y axis) and tensor modeling based on the proposed method HOSVD (Higher Order Singular Value Decomposition) + PCA + EFM using the multilinear algebra.

Keywords : Face verification, Depth image, IPC segmentation, Multi-blocs Local Descriptors, Dimensionality Reduction, Tensor representation.

الملخص

على مدى العقدين الماضيين، أصبح التعرف على الوجوه الآلي الأوتوماتيكي واحد من المهام الأكثر إثارة للاهتمام في مجال الرؤية الحاسوبية الذي يجذب الكثير من الاهتمام نظرا لتطبيقاته المختلفة. نظام التعرف على الوجه هو نهج بيومتري يستخدم الأساليب والتقنيات الآلية للتحقق من هوية الأشخاص استنادا إلى خصائص وجه الإنسان. خصصت هذه الأطروحة لنظام التحقق الآلي من الوجوه اعتمادا على صور العمق ثلاثية الأبعاد. للأسف، فإن الظروف غير المتحكم فيها مثل الاختلافات في الإضاءة، تعبيرات الوجه و دوران الرأس في عدة اتجاهات يؤثر سلبا على أداء هذه الأنظمة التي تعتمد على المعلومات ثنائية الأبعاد D2. لذلك، الحصول على أداء عالي لهذه الأنظمة في العالم الحقيقي يعتبر إشكالية مفتوحة. في الأونة الأخيرة، تحولت الأبحاث نحو التعرف على الوجه باستخدام المعلومات ثلاثية الأبعاد D3، والتي تمثل الخصائص الأكثر تمييزا. في البداية، أجرينا دراسة شاملة فيما يتعلق بأحدث تقنيات التعرف على الوجه D3 والخوارزميات التي واجهت تحديات التغييرات في الإضاءة، تعبيرات الوجه و مختلف زوايا الدوران. في هذه الأطروحة، تم اقتراح طريقة جديدة لتحديد موضع الوجوه ; تقسيم منطقة الوجه بواسطة تكامل منحنيات الإسقاط (IPC) تليها توطين طرف الأنف باستخدام صور العمق D3 والكثافة. النتائج التي تم الحصول عليها في قواعد البيانات CASIA3DV4 و GavabDB والمقارنة بينها وبين النتائج المتحصل عليها من طرف الباحثين الآخرين أكدت أهمية تحديد منطقة الوجه للحصول على دقة عالية في عملية التعرف عليها. بعد ذلك، فإننا نقترح إطار و طريقة جديدة للتحقق من الوجه D3 باستخدام تخفيض الأبعاد استنادا إلى المدرج الإحصائي للمواصفات المحلية متعددة الكتل لتمثيل صور الوجه. لهذا الغرض، يتم تقييم خمسة واصفات المحلية المختلفة LPQ,BSIF, FPLBP, TPLBP, LBP كما يتم تقييم أداء النهج المقترح مع مختلف خوارزميات تخفيض الأبعاد: PCA, PCA+EFM, OLPP. يستخدم المصنف متعدد الطبقات (SVM) لتنفيذ التحقق بين المحتالين والزبائن. في نهاية هذه الأطروحة، نقترح استراتيجية جديدة للتحقق من الوجه ثلاثي الأبعاد، هذه الإستراتيجية قوية وفعالة للتغيرات الكبيرة في زوايا الرأس، عن طريق التعلم بزوايا يولر (± 60 درجة على طول المحور Y) و النمذجة بواسطة الموتر على أساس النهج المقترح HOSVD+PCA+EFM باستخدام الجبر الخطي المتعدد.

الكلمات المفتاحية : التعرف على الوجوه, صور العمق, تقسيم منطقة الوجه بواسطة تكامل منحنيات الإسقاط, المواصفات المحلية متعددة الكتل, تخفيض الأبعاد, التمثيل بواسطة الموترات.

Table des matières

Dédicaces	ii
Remerciements	iii
Publications & Communications associées à la thèse	iv
Résumé	vi
Abstract	vii
الملخص	ix
Table des matières	x
Liste des figures	xiii
Liste des tableaux	xvi
Abréviations	xvii
Notations mathématiques	xviii

Chapitre 1 Introduction et Description Générale

1.1 Introduction générale.....	2
1.1.1 Contexte et motivation.....	2
1.1.2 Problématique	3
1.1.3 Objectifs et Contributions.....	5
1.1.4 Organisation de la thèse.....	6
1.2 La biométrie et les défis des systèmes de reconnaissance de visage	9
1.2.1 La biométrie et les différentes modalités.....	9
1.2.2 Description et conception du système de reconnaissance de visage (SRV).....	11
1.2.3 Défis scientifiques de la reconnaissance automatique de visage.....	13
1.2.3.1 La variation d'illuminations	13
1.2.3.2 La variation d'expressions faciales	14
1.2.3.3 La variation de poses.....	14
1.2.3.4 Autres difficultés.....	16
1.2.4 Principales bases de données	16
1.2.5 Évaluation et mesure de la performance.....	17
1.2.5.1 Mesure des performances pour le mode vérification	17
1.2.5.2 Mesure des performances pour le mode identification	19
1.3 Conclusion.....	19

Chapitre 2 État de l'Art sur la détection et la RV3D

2.1 Introduction	21
2.2 De la reconnaissance de visage 2D vers le 3D	21
2.3 La détection de visage	22
2.3.1 La détection de visage en 2D.....	23
2.3.2 La détection de visage en 3D basée sur le bout du nez	24
2.4 Les techniques d'acquisition de données du visage 3D.....	28
2.4.1 Les techniques actives	28

2.4.1.1 Le scanner laser 3D.....	28
2.4.1.2 Le scanner 3D à lumière structurée.....	28
2.4.2 La technique passive.....	29
2.5 La représentation de visage 3D	29
2.5.1 La représentation par nuage de points	29
2.5.2 La représentation par le maillage 3D (3D Mesh)	30
2.5.3 La représentation par profondeur.....	30
2.6 Les différentes approches de la reconnaissance de visage 3D	31
2.6.1 Approches d’alignement des surfaces 3D	32
2.6.2 Approches basées sur le modèle 3D	33
2.6.3 Approches holistiques 3D.....	35
2.6.4 Approches Géométriques.....	36
2.6.5 Approches 2D+3D	38
2.7 Conclusion.....	40

Chapitre 3 Reconnaissance 2D et 3D de Visage à base d’IPC

3.1 Introduction	42
3.2 Description de la méthode proposée	42
3.3 L'alignement du visage et l'enregistrement	43
3.4 La segmentation de la zone du visage et le prétraitement.....	44
3.4.1 Les images de profondeur et d'intensité	44
3.4.2 La segmentation de la zone du visage basée sur IPC	45
3.4.3 Prétraitements	46
3.5 La méthode combinée de réduction de dimension PCA+EFM.....	48
3.6 Classification.....	50
3.6.1 Classification par mesure de distance.....	50
3.6.2 Classification par SVM	50
3.7 Résultats expérimentaux	51
3.7.1 Résultats de l'étape de détection	52
3.7.2 Résultats du système de reconnaissance.....	55
3.8 Conclusion.....	60

Chapitre 4 Descripteurs Locaux et Réduction de Dimension pour Vérification V3D Expressions

4.1 Introduction	62
4.2 Travaux antérieurs.....	62
4.3 Aperçu de l'approche proposée	65
4.4 Les descripteurs locaux	66
4.4.1 Local Phase Quantization (LPQ).....	67
4.4.2 Local Binary Pattern (LBP)	69
4.4.3 LBP basé sur les patches (Patch based LBPs).....	70
4.4.3.1 Three-Patch Local Binary Patterns (TPLBP).....	70
4.4.3.2 Four-Patch Local Binary Patterns (FPLBP)	72

4.4.5 Binarized Statistical Image Features (BSIF)	73
4.5 Extraction des caractéristiques d'histogrammes	75
4.6 Réduction de dimension basée sur OLPP	76
4.7 Résultats expérimentaux et discussion	78
4.7.1 Base de données et protocole de vérification	78
4.7.2 Analyse de résultats et discussion	79
4.8 Conclusion.....	84

Chapitre 5 Vérification de V3D Poses /Angles d'Euler et Tenseurs

5.1 Introduction	86
5.2 Travaux de recherche autour de la vérification V3D poses	86
5.3 Vérification du visage basée sur les angles d'Euler	89
5.3.1 Aperçu de l'approche proposée	89
5.3.2 Génération de différentes poses avec la transformation d'Euler	90
5.3.3 Résultats et discussion	92
5.4 Vérification 3D de visage basée sur les tenseurs	95
5.4.1 Propriétés des tenseurs et modélisation multidimensionnelle	97
5.4.1.1 Définition d'un tenseur	97
5.4.1.2 Déploiement des tenseurs.....	97
5.4.1.3 Produit scalaire tensoriel.....	98
5.4.1.4 Produit extérieur.....	99
5.4.1.5 Opérateur n -mode produit.....	99
5.4.2 Décomposition tensorielle par HOSVD	99
5.4.3 Description de l'approche proposée	101
5.4.4 Résultats et discussion	103
5.5 Conclusion.....	107

Chapitre 6 Conclusion et Perspectives

6.1 Conclusion générale	110
6.2 Perspectives	110
Bibliographie	116

Liste des figures

Chapitre 1 Introduction et Description Générale

Figure 1.1 Les différents types de modalités biométriques.....	11
Figure 1.2 Comparaison des différentes caractéristiques biométriques [20]	11
Figure 1.3 Schéma de fonctionnement d'un système biométrique. Diagrammes des processus : apprentissage, vérification et identification.	12
Figure 1.4 Exemple de variation d'illuminations pour différentes personnes.....	14
Figure 1.5 Exemple de quelques variations d'expressions faciales de la même personne.....	14
Figure 1.6 Un exemple d'angles de rotation : Pitch, Yaw et Roll.....	15
Figure 1.7 Un exemple des applications pour l'estimation de la pose de la tête ; analyse d'émotion d'un conducteur de voiture ; la vidéosurveillance	15
Figure 1.8 Exemples de variations de poses dans la base de données CASIA3DV4.....	16
Figure 1.9 Exemples du vieillissement et de l'occlusion du visage.....	16
Figure 1.10 Exemples des courbes ROC; (a) TV Vs TFA, (b) TFR Vs TFA	18
Figure 1.11 Exemple de la courbe CMC	19

Chapitre 2 État de l'Art sur la Détection et la RV 3D

Figure 2.1 Exemple de la détection de bout du nez avec différentes poses [7].....	25
Figure 2.2 L'approche proposée par Anuar L.H et al pour la détection de bout du nez [29] ..	25
Figure 2.3 Le système de filtrage hiérarchique pour la détection de nez [28]	26
Figure 2.4 Illustration de la méthode de Mian; (a) la détection de la pointe du nez, (b) sphère centrée sur le bout du nez d'un nuage de points 3D pour recadrer le visage [38].....	26
Figure 2.5 Les scanners 3D les plus populaires; (a) Minolta Vivid 910, (b) Cyberware 3030, (c) Microsoft Kinect.....	29
Figure 2.6 Représentation par nuage de points pour un visage CASIA3DV4	30
Figure 2.7 Représentation par maillage 3D pour un visage GavabDB [8].....	30
Figure 2.8 Représentation de visage par profondeur; (a) système de coordonnées 3D, (b) image de profondeur dans le plan (X,Y)	31
Figure 2.9 Alignement de deux visages 3D [2]	32
Figure 2.10 Le modèle 3DMM proposé par Blenz et al [67]	34
Figure 2.11 L'organigramme du modèle PAF proposé par [69].....	34
Figure 2.12 Illustration de l'approche proposée par Al-Osaimi et al [77]	35
Figure 2.13 Extraction des courbes nasales [91]	37
Figure 2.14 Vue d'ensemble de la méthode proposée par Xu et al [14].....	39
Figure 2.15 Classification des approches de la reconnaissance de visage 3D	39

Chapitre 3 Reconnaissance 2D et 3D de Visage à base d'IPC

Figure 3.1 Organigramme de notre système de reconnaissance.....	43
Figure 3.2 Détection de visage basée sur IPC	44
Figure 3.3 L'information de visage dans CASIA3DV4; A: image de profondeur 3D, B: image d'intensité 2D (couleur), C: nuage de points 3D	45
Figure 3.4 Le processus de détection faciale basée sur IPC; (a) binarisation de l'image de profondeur, (b) projection intégrante horizontale, (c) projection intégrale verticale, (d) zone du visage segmentée	46
Figure 3.5 Détection et prétraitement de l'image de profondeur; (a) image de profondeur d'entrée, (b) segmentation de la zone du visage par IPCs, (c) élimination du bruit et remplissage des trous (d), image de profondeur après masque elliptique	47

Figure 3.6 Détection et prétraitement de l'image intensité; (a) image intensité d'entrée, (b) segmentation de la zone du visage par l'IPCs, (c) élimination de bruit et remplissage des trous, (d) image de profondeur après l'égalisation d'histogramme et masque elliptique.	47
Figure 3.7 Classification SVM; (a) SVM linéaire séparation par une ligne droite, (b) SVM non linéaire séparation par une courbe	51
Figure 3.8 Les variations d'expression dans la base de données CASIA3DV4	53
Figure 3.9 Les variations d'illuminations dans la base de données CASIA3DV4	53
Figure 3.10 Les variations de poses dans la base de données CASIA3DV4	53
Figure 3.11 Exemples des 9 vues pour la même personne dans GavabDB	54
Figure 3.12 Variations de TEE en fonction du nombre de caractéristiques	58
Figure 3.13 Variations de TV en fonction du nombre de caractéristiques	58

Chapitre 4 Descripteurs Locaux et Réduction de Dimensionnalité pour Vérification V3D Expressions

Figure 4.1 Organigramme de l'approche proposée	66
Figure 4.2 Prétraitement de l'image de profondeur 3D; (a) image d'entrée, (b) image détectée, (c) image prétraitée	66
Figure 4.3 Organigramme de l'ensemble des étapes nécessaires du descripteur LPQ	68
Figure 4.4 Représentation d'une image de profondeur avec le descripteur LPQ sous différents voisinage de pixel N_z ; (a) image d'entrée, (b) code LPQ ($N_z=3$), (c) code LPQ ($N_z=5$), (d) code LPQ ($N_z=7$), (e) code LPQ ($N_z=9$)	68
Figure 4.5 Une illustration de LBP basique	69
Figure 4.6 Exemples de d'opérateur LBPP. R; (a) LBP8.1, (b) LBP8.2, (c) LBP16.2	69
Figure 4.7 Représentation d'une image de profondeur avec le descripteur LBP	70
Figure 4.8 La procédure de descripteur TPLBP	71
Figure 4.9 La représentation TPLBP d'une image de profondeur avec différentes valeurs de rayons et de tailles des patches w	71
Figure 4.10 La procédure du descripteur FPLBP	72
Figure 4.11 La représentation FPLBP d'une image de profondeur avec différentes valeurs de tailles de patches w et de nombre des patches voisins S	73
Figure 4.12 Les 13 images naturelles utilisées pour l'apprentissage des filtres dans le descripteur BSIF [97]	74
Figure 4.13 Filtres de taille 9×9 ($l = 9$)	74
Figure 4.14 La représentation BSIF d'une image de profondeur avec différentes tailles (l) du filtre et différentes longueur de la chaîne de bits n	75
Figure 4.15 Les descripteurs et l'extraction de caractéristiques d'histogramme	76
Figure 4.16 Image de profondeur de la même personne, première ligne variations d'illuminations, deuxième ligne variation d'expressions, troisième ligne variations d'expressions sous différentes illuminations	78
Figure 4.17 Représentation du visage basée sur les descripteurs locaux; (a) image d'entrée; (b) code image LPQ; (c) code image TPLBP, (d) code image FPLBP, (e) code image BSIF, (f) code image LBP	79
Figure 4.18 Courbes ROCs avec PCA	81
Figure 4.19 Courbes ROCs avec OLPP	81
Figure 4.20 Courbes ROCs avec PCA+EFM	82
Figure 4.21 Variation de taux de vérification en fonction du nombre de caractéristiques	83

Chapitre 5 Vérification de V3D Poses /Angles d'Euler et Tenseurs

Figure 5.1 Résolution du problème de données manquantes utilisant la symétrie du visage sur les zones occultées [153].....	87
Figure 5.2 Normalisation de visage 3D proposée par <i>Mian et al</i> [38]	88
Figure 5.3 Signatures de profil tourné selon Y; les images du haut montrent les modèles 3D avec le profil extrait et les images du bas montrent les profils extraits associés [155]	89
Figure 5.4 Estimation de la pose basée sur RRF [157]	89
Figure 5.5 Organigramme de la méthode de vérification de visage 3D proposée	90
Figure 5.6 Les angles de rotation de la tête β_i générés par la transformation d'Euler pour une image de profondeur	92
Figure 5.7 Variation de taux de vérification en fonction du rayon et du nombre des blocs LPQ	95
Figure 5.8 Courbes ROCs pour différents nombres de blocs.....	95
Figure 5.9 Représentation tensorielle d'ensemble d'images.....	96
Figure 5.10 Représentation matérielle (vectorisation) d'ensemble d'images	96
Figure 5.11 Déploiement d'un tenseur d'ordre 3 dans les différents modes.....	98
Figure 5.12 Décomposition de Tucker3 d'un tenseur d'ordre 3	100
Figure 5.13 Représentation du tenseur d'ordre 3 proposé.....	101
Figure 5.14 Présentation générale de l'approche proposée	102
Figure 5.15 Performances de vérification avec HOSVD; (a) TEE en fonction de la dimension de tenseur D1, (b) TV en fonction de la dimension de tenseur D1	105
Figure 5.16 Performances de vérification avec PCA+EFM et HOSVD+PCA+EFM; (a) TEE en fonction du nombre de caractéristiques, (b) TV en fonction du nombre de caractéristiques.....	106
Figure 5.17 Courbes ROCs des trois méthodes.....	107

Liste des tableaux

Chapitre 1 Introduction et description générale

Tableau 1.1 Les BDD 2D et 3D proéminentes et leurs caractéristiques.....	17
--	----

Chapitre 3 Reconnaissance 2D et 3D de visage à base d'IPC

Tableau 3.1 La précision de la détection dans CASIA3DV4	54
Tableau 3.2 La précision de la détection dans GavabDB.....	54
Tableau 3.3 Les protocoles utilisés pour la base de données CASIA3DV4.....	56
Tableau 3.4 Le protocole utilisé pour la base de données GavabDB	57
Tableau 3.5 Meilleurs résultats de reconnaissance pour différentes expériences	59
Tableau 3.6 Comparaison de TV sur GavabDB	59
Tableau 3.7 Comparaison de TV sur CASIA3DV4	59

Chapitre 4 Descripteurs Locaux et Réduction de Dimensionnalité pour Vérification V3D Expressions

Tableau 4.1 Protocole utilisé pour la vérification.....	78
Tableau 4.2 Paramètres optimaux des descripteurs utilisés	79
Tableau 4.3 Performances de la vérification avec PCA durant l'évaluation et le test	80
Tableau 4.4 Performances de la vérification avec OLPP durant l'évaluation et le test	80
Tableau 4.5 Performances de la vérification avec PCA+EFM durant l'évaluation et le test... 80	
Tableau 4.6 Performances de la vérification utilisant PCA+EFM basé sur la combinaison de LPQ avec LBP, FPLBP et BSIF	82
Tableau 4.7 Comparaison des taux de vérification avec l'état de l'art.....	84

Chapitre 5 Vérification de V3D Poses /Angles d'Euler et Tenseurs

Tableau 5.1 Protocole utilisé pour l'évaluation de variations de poses	93
Tableau 5.2 Performances de la vérification, effet du rayon de LPQ	93
Tableau 5.3 Performances de la vérification, effet du nombre des blocs.....	94
Tableau 5.4 Exemples de modélisations tensorielles	97
Tableau 5.5 L'algorithme de décomposition en HOSVD	101
Tableau 5.6 Protocole utilisé pour l'évaluation des trois méthodes.....	103
Tableau 5.7 Performances de la vérification avec PCA+EFM	104
Tableau 5.8 Performances de la vérification avec HOSVD.....	105
Tableau 5.9 Performances de la vérification avec HOSVD+PCA+EFM	105
Tableau 5.10 Comparaison des performances de différentes méthodes	106

Abréviations

BSIF: Binarized Statistical Image Features.

BDD: Base De Données.

CMC: Cumulative Match Characteristic.

EFM: Enhanced Fisher linear discriminant Model.

FPLBP: Four-Patch Local Binary Patterns.

HOSVD: Higher Order Singular Value Decomposition.

HTER: Half Total Error Rate.

ICP: Iterative Closest Point.

IHM: Interactions Homme Machine.

IPC: Integral Projection Curves.

LDA: Linear Discriminant Analysis.

LBP: Local Binary Patterns.

LPQ: Local Phase Quantization.

OLPP: Orthogonal Locality Preserving Projection.

PIN: Personal Identification Number.

PCA: Principal Component Analysis.

ROC: Receiver Operating Characteristic.

R1RR: Rank-1 Recognition Rate.

SIFT: Scale-Invariant Feature Transform.

SVM: Machines à vecteurs de support.

SRV: Système de Reconnaissance de Visage.

TFA: Taux de Fausse Acceptation.

TFR: Taux de Faux Rejet.

TEE: Taux d'Égale Erreur.

TV: Taux de vérification.

TPLBP: Three-Patch Local Binary Patterns.

V3D: Visage 3D.

3DRM: 3D Rotation Matrix.

Notations mathématiques

P_i : Point dans le nuage des points 3D.

$HP(\mathbf{y})$: La courbe de projection horizontale.

$VP(\mathbf{x})$: La courbe de projection verticale.

\mathbf{A} : Matrice d'apprentissage.

M : Le nombre des images d'apprentissage.

N : La taille du vecteur caractéristique.

$\bar{\mathbf{A}}$: La moyenne de chaque visage.

$\bar{\mathbf{m}}$: La moyenne de toutes les images d'apprentissage.

\mathbf{C} : La matrice de covariance.

\mathbf{U}_{PCA} : Matrice de transformation linéaire PCA.

\mathbf{W}_i : Le vecteur caractéristique d'apprentissage projeté dans le sous-espace de vecteurs propres.

\mathbf{S}_w : Matrices de dispersion inter-classes.

\mathbf{S}_b : Matrices de dispersion inter-classes.

\mathbf{U}_{EFM} : La matrice de transformation EFM.

\mathbf{K}_b : La nouvelle matrice de dispersion inter-classes.

\mathbf{W}_{final} : La matrice d'apprentissage finale.

u : Fréquence.

$S(\mathbf{A}, \mathbf{B})$: La corrélation entre deux vecteurs A et B .

\mathbf{w}_u : Les vecteurs de base de la décomposition à la fréquence u .

$q_j(\mathbf{x})$: Quantificateur scalaire.

F_{LPQ} : L'étiquette d'image du descripteur LPQ.

I_c : Pixel central de l'image d'entrée.

S_i : Filtre de descripteur BSIF.

l : Taille du filtre de descripteur BSIF.

H : L'histogramme de l'image faciale.

\mathbf{W}_{OLPP} : La matrice de transformation OLPP.

$(\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i)$: Le nuage de points d'entrée.

$(\mathbf{x}_i^\beta, \mathbf{y}_i^\beta, \mathbf{z}_i^\beta)$: Le nuage de points tourné avec angle β .

A : Tenseur d'ordre N .

\mathbf{A} : Matrice de déploiement du tenseur A .

Chapitre 1

Introduction et Description Générale

1.1 Introduction générale.....	2
1.1.1 Contexte et motivation.....	2
1.1.2 Problématique.....	3
1.1.3 Objectifs et Contributions.....	5
1.1.4 Organisation de la thèse.....	6
1.2 La biométrie et les défis des systèmes de reconnaissance de visage	9
1.2.1 La biométrie et les différentes modalités.....	9
1.2.2 Description et conception du système de reconnaissance de visage (SRV).....	11
1.2.3 Défis scientifiques de la reconnaissance automatique de visage.....	13
1.2.3.1 La variation d'illuminations	13
1.2.3.2 La variation d'expressions faciales.....	14
1.2.3.3 La variation de poses.....	14
1.2.3.4 Autres difficultés.....	16
1.2.4 Principales bases de données	16
1.2.5 Évaluation et mesure de la performance.....	17
1.2.5.1 Mesure des performances pour le mode vérification.....	17
1.2.5.2 Mesure des performances pour le mode identification	19

1.1 Introduction générale

1.1.1 Contexte et motivation

Récemment la reconnaissance automatique des visages est devenue l'une des tâches les plus intéressantes et importantes dans le domaine de la vision par ordinateur et la biométrie, concernant les méthodes théoriques et les logiciels afin de reconnaître les personnes en fonction de leurs images de visage. La reconnaissance du visage, comme un sujet de recherche interdisciplinaire, possède une longue histoire de l'étude par la communauté scientifique de la recherche depuis les années 80 [1]. Après cette période, la reconnaissance automatique de visage a parcouru un long chemin et le passage vers des produits commerciaux n'a reçu une grande impulsion qu'à partir de l'année 1994 en grande partie grâce à la mise en œuvre d'un programme d'évaluation internationale FERET (Face Recognition Technology) [2].

L'importance de la reconnaissance de visage est due aux nombreuses applications qui peuvent être rencontrées dans différents domaines tels : la sécurité (le contrôle d'accès aux zones autorisées, les ordinateurs, les aéroports, l'identification des individus, etc ...), les dispositifs de surveillance dans les espaces publics (stades, gares, grands centres commerciaux, etc...), les applications médico-légales (vérification de l'identité pour le système de la justice pénale, l'identification des victimes de catastrophes), l'identité des personnes dans les bases de données image/vidéo, l'interface homme-machine, les cartes à puce (renforcement de la sécurité des guichets automatiques (Automated Teller Machine) et le passeport biométrique aussi connu sous le nom de Epassport) [3]. Le domaine de la reconnaissance faciale exploite les connaissances de nombreuses disciplines telles que : le traitement d'images, les mathématiques appliquées, la reconnaissance des formes, l'apprentissage automatique, la perception visuelle, la psychophysique et les neurosciences.

De plus en plus, les applications en traitement d'image comportent des données multidimensionnelles. Cependant, la majorité des algorithmes de traitement sont basés sur des modèles de données vectoriels ou matriciels. Ces opérations détruisent la structure des données et peuvent entraîner une baisse des performances et/ou de la robustesse des traitements dans diverses applications. Il est donc nécessaire d'adapter les traitements classiques pour ces nouvelles configurations (multidimensionnelles). L'algèbre multilinéaire [4, 5] permet d'exploiter ces données tout en conservant leur structure. Les données sont alors représentées sous forme de tableaux multidimensionnels appelés **tenseurs**. Plus récemment, ces techniques ont été aussi étendues aux données de la biométrie faciale. Cependant, la généralisation des traitements de données faciales sous forme vectoriels ou matriciels classiques au cas tensoriel

n'est pas encore évidente. Nous envisageons le développement d'un système de vérification de visage basé sur les tenseurs et les techniques d'algèbre multilinéaire afin d'aborder les défis d'illuminations spécialement d'expressions et de grandes variations de poses.

Généralement un système de reconnaissance de visage se compose de deux phases, la phase d'apprentissage (off-line) et la phase de test (on-line). La phase d'apprentissage sera effectuée une seule fois dans laquelle l'enrôlement des images faciales des différents individus est utilisé afin d'extraire la signature biométrique de chaque individu. Ces données d'apprentissage sont préparées pour le classificateur afin de faire la reconnaissance. Au cours de la phase de test, les nouvelles données sont classifiées avec les données d'entraînement qui sont apprises dans la phase d'apprentissage. Le même traitement est effectué dans les deux phases. Pour atteindre cet objectif, généralement la procédure de traitement d'un système de reconnaissance de visage est subdivisée en trois étapes principales [6] : *la détection des visages, l'extraction de caractéristiques et la reconnaissance (classification)*.

1.1.2 Problématique

L'obtention des hautes performances pour un système de reconnaissance de visage dans le monde réel est un problème ouvert aux chercheurs depuis quelques années. Malheureusement, les conditions non contrôlées telles que les variations d'illumination, les occultations, les expressions faciales et les variations de poses affectent considérablement les performances des systèmes de reconnaissance faciale surtout celles basées sur l'information 2D, car ce genre d'informations dépend principalement des sources de la lumière, en plus l'image 2D ou bien l'image couleur ne représente pas la forme du visage et ne traite pas le visage comme un objet, alors que ces systèmes sont sensibles à l'environnement réel non contrôlé.

La variation de poses de la tête est un problème major pour la reconnaissance de visage [7]. La correction de la pose et l'estimation de l'angle de rotation de la tête sont des processus nécessaires pour résoudre ces problèmes. Cependant, à cause de la complexité mathématique et pour des raisons de coûts élevés en termes de mémoire et de temps de calcul, le développement d'un système de reconnaissance automatique du visage robuste à la variation de poses est considéré comme un grand défi pour les chercheurs de la biométrie faciale.

Autre difficulté pour les systèmes de reconnaissance automatique de visage la représentation des images faciales et l'extraction des caractéristiques les plus discriminantes pour construire la signature biométrique des personnes. Dans notre travail, cette signature se présente sous forme d'un vecteur caractéristique. Plusieurs méthodes ont été développées pour l'extraction des caractéristiques qui sont des méthodes globales ou locales. Récemment, les

descripteurs locaux sont utilisés comme des méthodes très efficaces pour la représentation des images, ils sont proposés à l'origine pour la classification des textures. Ensuite, ils ont été utilisés pour l'analyse des images faciales. Dans ce mémoire de thèse, nous étudions et évaluons cinq parmi les meilleurs descripteurs utilisés dernièrement dans la littérature.

Plus récemment, la mise au point de la recherche en biométrie faciale s'est orientée vers l'utilisation de la surface faciale 3D et la forme du visage pour la représentation des caractéristiques les plus discriminantes avec une augmentation de la dimensionnalité des vecteurs caractéristiques [8]. Pour réduire l'impact de ce genre de problèmes, la réduction de dimension est utilisée dans la plupart des systèmes de reconnaissance faciale. La réduction de dimensionnalité joue un rôle important dans le domaine du traitement de l'information notamment dans les systèmes de la biométrie faciale qui nécessitent une rapidité d'exécution et une grande précision. Plusieurs approches ont été proposées dans la littérature, notons que les deux algorithmes, PCA (Principal Component Analysis) et LDA (Linear Discriminant Analysis) sont des algorithmes de base pour la plupart des approches proposées [9]. De nombreux travaux de recherche concernant la reconnaissance automatique de visages basés sur les images 3D ont été proposés [10-14]. Dans notre travail, nous utilisons l'image de profondeur qui représente les caractéristiques géométriques des visages humains 3D, en raison de leurs nombreux avantages par rapport aux images d'intensité 2D pour réduire l'effet des variations d'éclairage, d'expressions et de poses [10, 11].

La problématique principale de notre travail de recherche concerne la présence des conditions non contrôlées dans le système de reconnaissance automatique de visage : les variations d'illuminations, d'expressions faciales et des poses. Ainsi que l'extraction des caractéristiques discriminantes des images faciales basées sur les descripteurs locaux en utilisant les images de profondeur. Toutefois, nous examinons le rôle du processus de la réduction de données pour développer un système de reconnaissance automatique de visage robuste, efficace et précis.

Par la suite, nous examinons la contribution de la représentation sous forme de tenseurs des données faciales basée sur l'algèbre multilinéaire pour modéliser les différents défis telles que l'illumination, les expressions et les poses afin de développer un système de vérification de visage performant et robuste à ces problèmes. La rotation du visage dans un repère 3D, la décomposition des tenseurs et la projection des données d'apprentissage dans le sous-espace multilinéaire sont des questions que nous allons contribuer à résoudre dans ce travail de thèse.

1.1.3 Objectifs et Contributions

L'objectif principal de la première étape de notre système, est la détection et la localisation de la zone faciale dans une image donnée. Après l'étape de détection, les signatures biométriques de visage humain sont extraites sous forme d'un vecteur caractéristique au sein de la deuxième étape. Après avoir formulé la représentation de chaque image de visage, la dernière étape consiste à reconnaître les individus. Pour chaque personne, plusieurs images sont prises et leurs caractéristiques sont extraites et stockées dans la base de données (phase off-line). Puis, quand une image de visage d'entrée arrive, nous effectuons la détection de visage, l'extraction de caractéristiques et enfin la comparaison de ces caractéristiques à chaque classe de visage stocké dans la base de données (phase on-line). L'étape de reconnaissance se fait en deux modes : *Identification* ou bien *Vérification*. Ces deux modes sont abordées plus en détail dans la section 1.2.2. Nous nous intéressons dans tous nos expériences au mode vérification.

L'objectif de la recherche de ce mémoire de thèse est de développer, implémenter et évaluer un nouveau système automatique de reconnaissance de visage basé sur les images de profondeur, dans lesquelles les variations : d'illuminations, d'expressions et de poses sont très différentes entre l'ensemble d'apprentissage et l'ensemble de test. Nous proposons en particulier une nouvelle méthode de détection du visage. D'autre part, une stratégie efficace de représentation des caractéristiques locales basée sur les histogrammes des descripteurs locaux multi-blocs et la réduction de la dimensionnalité sont proposées et évaluées. Un chapitre est consacré à la mise en œuvre d'un SRV pour résoudre le problème de reconnaissance en présence de poses en utilisant l'algèbre multilinéaire.

Un bon système de reconnaissance doit être :

- **Automatique** : sans aucune intervention manuelle, dans toutes les étapes de traitement en cours de la phase d'apprentissage ou bien la phase de test.
- **Précis** : le système devrait présenter un taux global de reconnaissance au-dessus de 90% avec des taux d'erreurs faibles (voir la section 1.2.5).
- **Robuste** : le système devrait fonctionner avec succès dans de différentes conditions d'illuminations, d'expressions faciales et de variations de poses.

Les contributions principales de notre travail peuvent être résumées comme suit :

- Etude de l'état de l'art des techniques récentes basiques sur la détection et la reconnaissance de visage 3D.

- Développement d'une méthode de détection de visage (IPC based facial area segmentation) basée sur la segmentation de la zone faciale en utilisant les courbes verticales et horizontales de projection d'intégrales [9, 15, 16].
- Évaluation d'un nouveau système de vérification de visage 3D basé sur les histogrammes de cinq descripteurs locaux multi blocs : LBP, TPLBP, FPLBP, BSIF et LPQ. Nous donnons une analyse détaillée de ces algorithmes et nous comparons leurs performances pour construire un système automatique et efficace de vérification 3D de visage. Dans laquelle nous pourrions résoudre les défis où les expressions faciales, les conditions d'illumination et les variations de poses dans les données d'apprentissage et de test sont très différents.
- Etude de l'efficacité de l'extraction des caractéristiques faciales locales dans des petits rectangles basés sur les histogrammes en se basant sur les travaux [17, 18].
- Etude et conception de la fusion par combinaison du descripteur LPQ avec différents descripteurs au niveau des caractéristiques par la concaténation des histogrammes.
- Évaluation et comparaison des trois techniques de réduction de dimension : PCA, OLPP et PCA + EFM [17, 18].
- Proposition d'une vérification de visage 3D en présence d'une grande variation de poses en utilisant l'apprentissage du système avec les angles d'Euler (poses : $\pm 60^\circ$ selon l'axe Y).
- Proposition d'une nouvelle stratégie de vérification de visage 3D robuste et efficace aux grandes variations de poses en utilisant : l'apprentissage par les angles d'Euler et la modélisation tensorielle basée sur une méthode proposée qui se compose de deux étapes de réduction de dimension multilinéaire HOSVD et linéaire PCA+EFM.
- Enfin, proposition et validation d'un modèle basé sur les angles d'Euler et un autre basé sur l'utilisation des tenseurs pour remédier aux problèmes rencontrés dans la reconnaissance de visage en présence de poses.

1.1.4 Organisation de la thèse

Le mémoire de la thèse s'articule autour de six chapitres :

Le **premier chapitre** est une partie introductive au contexte de cette thèse consacré à la présentation de la problématique, les objectifs et les contributions de notre travail. Une description générale de la biométrie et la reconnaissance automatique de visage sont décrites. Dans lesquelles, nous présentons les différentes modalités biométriques ainsi que la place du visage parmi celles-ci. Ensuite, nous introduisons la structure générale et le principe de fonctionnement d'un système de reconnaissance de visage, ses différents défis scientifiques

dans les environnements non contrôlés et les paramètres d'évaluation ainsi que la mesure des performances de ces systèmes. Par ailleurs, nous mettons également en évidence les principales bases de données 2D et 3D qui sont largement utilisées dans la littérature.

Le **deuxième chapitre** est consacré à l'état de l'art de la détection et la reconnaissance de visage 3D. Tout d'abord, nous présentons en revue des méthodes de détection de visage 2D et 3D. Plus spécifiquement, nous abordons les méthodes de la détection 3D basées sur le bout du nez comme un point de référence. Ensuite, nous présentons les techniques d'acquisition de données faciales 3D et les différents types de données pour représenter un visage humain sous forme tridimensionnelle. Enfin, nous présentons une classification des approches de la reconnaissance de visage 3D en cinq catégories : approches d'alignement des surfaces 3D, approches basées sur le modèle 3D, approches holistiques 3D, approches géométriques, et approches 2D+3D.

Le **troisième chapitre** est une description de la méthode de détection proposée qui consiste en la reconnaissance de visage 2D et 3D utilisant les courbes de projection intégrale basée sur les image de profondeur et d'intensité en utilisant la segmentation de la zone du visage avec les courbes de projection intégrale (IPC based Facial Area Segmentation) horizontale et verticale. Nous terminerons ce chapitre par la présentation et la discussion des résultats obtenus par notre système de reconnaissance basé sur la méthode de détection IPC en présence des variations d'illuminations, d'expressions faciales et de poses. Une comparaison de notre approche avec l'état de l'art est présentée à la fin de ce chapitre.

Dans le **quatrième chapitre** nous proposons une méthode d'évaluation des caractéristiques d'histogrammes des descripteurs locaux multi-blocs et de la réduction de dimensionnalité pour la vérification de visage 3D basé sur les images de profondeur en présence d'expressions et d'illuminations. Tout d'abord, une présentation générale de la méthode proposée est introduite. Nous passons en revue les travaux connexes des méthodes de vérification basée sur les descripteurs locaux et la réduction de dimensionnalité. Ensuite, nous mettons en avant une analyse mathématique des descripteurs locaux LBP, TPLBP, FPLBP, BSIF et LPQ utilisés dans notre travail ainsi que la méthode de réduction de données OLPP. Cette dernière est comparée avec les deux méthodes PCA et PCA+EFM. Une validation expérimentale de l'approche proposée comprenant l'évaluation de notre système et la discussion des résultats est donnée à la fin de ce chapitre.

Le **cinquième chapitre** est divisé en deux parties essentielles. Dans un premier temps, nous introduisons une nouvelle méthode de vérification de visage 3D robuste et efficace aux variations de poses basée sur l'apprentissage du système avec les angles d'Euler ($\pm 60^\circ$ selon

l'axe Y). Basés sur les résultats obtenus dans le chapitre 4, nous utilisons les images de profondeur avec les histogrammes de descripteur LPQ pour représenter chaque image de personne dans la base de données. Ensuite, PCA+EFM est utilisée pour réduire la dimension des vecteurs caractéristiques. Dans la deuxième partie, nous proposons une mise au point d'une nouvelle approche pour la vérification de visage en présence de grande variété de poses basé sur l'analyse tensorielle. Les défis : illuminations, expressions et poses sont considérés comme des problèmes d'algèbre multilinéaire dans laquelle l'ensemble des images est représenté comme un tenseur d'ordre 3. Ce tenseur est décomposé basé sur la méthode proposée HOSVD + PCA + EFM afin de réduire sa dimension. Une de nos nouvelles contributions dans ce chapitre est une généralisation multilinéaire de la méthode PCA+EFM.

Dans le **sixième chapitre**, nous concluons ce mémoire de thèse en résumant les points principaux de nos contributions et nous proposons des perspectives envisageables à explorer suite à notre travail.

1.2 La biométrie et les défis des systèmes de reconnaissance de visage

1.2.1 La biométrie et les différentes modalités

Le mot biométrie signifie «mesure biologique». Un système biométrique est un système de mesure automatique qui consiste à identifier un individu à partir de ses caractéristiques biologiques, comportementales ou physiologiques afin de faire l'authentification (vérification si l'identité d'une personne est vraie ou fausse) ou bien l'identification (reconnaître un individu parmi M individus). Il existe plusieurs modalités biométriques pour reconnaître les personnes. Dans la suite, nous listons les différentes modalités biométriques avec leurs avantages et inconvénients :

- **La séquence d'ADN** représente l'information génétique, elle est surtout utilisée dans les applications de médecine légale et n'est pas utile dans les applications de reconnaissance automatique en temps réel.
- **La signature** a été largement utilisée, mais elle pourrait être affectée par l'état physique et émotionnel de la personne. Elle peut être modifiée avec le temps et est sensible à la fraude.
- **L'empreinte digitale** est parmi les modalités biométriques les plus utilisées. Sa performance est très élevée avec un prix raisonnable, mais l'un de ses inconvénients est que les empreintes digitales d'une personne peuvent être endommagées par des coupures ou des brûlures rendant ainsi les données biométriques inutiles. La construction de ces systèmes, nécessite beaucoup de ressources de calcul.
- **La géométrie de la main** est en effet, celle utilisée dans les premiers systèmes biométriques. Elle est très simple à mettre en œuvre, facile à utiliser et relativement pas chère. Cependant, la géométrie de la main n'est pas très caractéristique. Elle a les mêmes inconvénients que l'empreinte digitale.
- **Les veines de la main** peuvent être capturées par des caméras infrarouges, l'acquisition d'image est assez difficile et la mise en place du système est assez coûteuse.
- La texture de **l'iris** fournit des informations très utiles et efficaces pour la reconnaissance et la discrimination des personnes. Ces systèmes étaient très chers, mais les nouveaux systèmes sont devenus plus économiques. Cependant, la collection des données de l'iris est un peu fastidieuse.
- **La rétine** est la modalité biométrique la plus sécurisée, car il n'est pas facile de la changer ou la répliquer. La rétine possède des caractéristiques uniques pour chaque individu et pour chaque œil de l'être humain. Cependant, l'acquisition des données de la rétine n'est pas facile. Par conséquent, l'acceptation par le public est un peu difficile.

- **La voix** est rarement utilisée, elle n'est pas très spécifique et change beaucoup au cours d'une période de temps. La voix n'est pas utile dans la reconnaissance des personnes à grande échelle.
- L'analyse de **la démarche** est l'étude systématique du mouvement humain. En effet, c'est une technique biométrique émergente, dans laquelle, les gens sont identifiés uniquement par l'analyse de leur démarche. Bien que la recherche soit toujours en cours, la démarche a suscité un intérêt en tant que méthode d'identification, car elle ne nécessite pas une coopération de la personne. Toutefois, la démarche peut être affectée par : la grossesse des femmes, après un accident, une maladie affectant les jambes, ou un après sévère gain de poids etc...
- **Le visage** est probablement l'une des caractéristiques biométriques la plus commune et la plus utilisée, car elle est naturelle, non intrusive, sécurisée et très acceptable par le public. Ce système est praticable et la collection de données peut être réalisée d'une manière simple avec différents types de caméras (caméras numériques simples, caméras infrarouges, scanners 3D).

Le visage, l'empreinte digitale, l'iris et la voix sont les modalités biométriques les plus populaires. Récemment, les progrès dans le domaine de la reconnaissance des personnes ont permis d'explorer d'autres modalités biométriques prometteuses telles que : la thermographie (thermo-gramme de la main et le thermo-gramme facial), les battements du cœur, l'odeur corporelle ou encore le mouvement des yeux [19]. L'avantage majeur de ces modalités c'est qu'elles ne sont pas intrusives. Malheureusement, ces solutions biométriques restent trop coûteuses. La figure 1.1 montre les différents types des modalités biométriques.

Le visage humain est une modalité biométrique très efficace. En plus d'être naturelle, non intrusive et précise, l'avantage le plus important du visage humain est qu'il peut être capturé à distance et de manière secrète. Parmi les six caractéristiques biométriques considérées dans [20] (le visage, les empreintes digitales, la géométrie de la main, la voix, les yeux et la signature) le visage possède le plus grand score de compatibilité (voir figure 1.2) sur la base du système de compatibilité MRTD (Machine Readable Travel Documents), selon un certain nombre de facteurs d'évaluation (l'inscription, le renouvellement des données, les exigences de vérification de l'identité assistée par machine, la redondance, la perception du public) [20]. Pour cela, un énorme nombre des travaux de recherche ont été réalisés au cours des trois dernières décennies. Cependant, l'obtention de haute performance pour un système de reconnaissance de visage dans le monde réel avec les scénarios les plus pratiques reste encore en recherche.



Figure 1.1 Les différents types de modalités biométriques

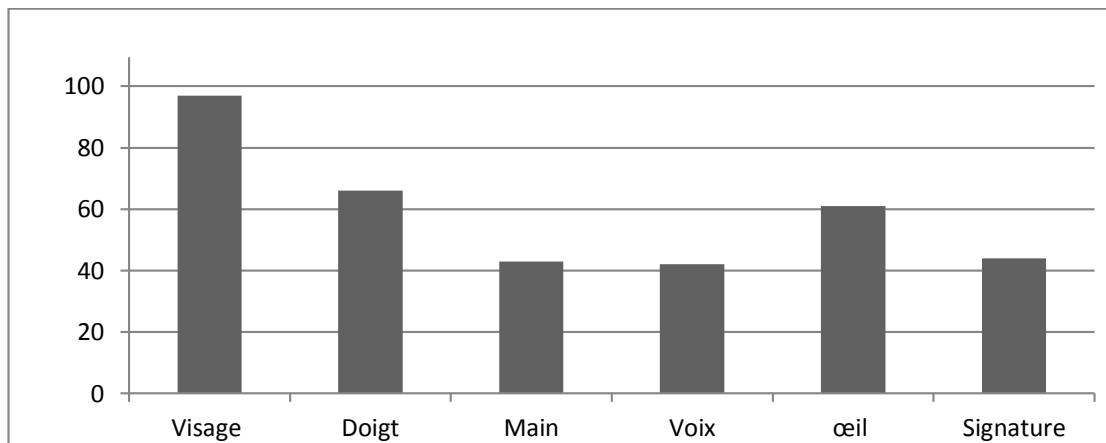


Figure 1.2 Comparaison des différentes caractéristiques biométriques [20]

1.2.2 Description et conception du système de reconnaissance de visage (SRV)

Le système de reconnaissance de visage selon le contexte de l'application utilisée (Identification/ Authentification) se compose de deux phases :

- L'enrôlement (l'apprentissage) : c'est une phase commune, on peut l'appeler aussi la phase off-line dans laquelle, la signature biométrique de chaque visage est extraite sous forme d'un vecteur de caractéristiques.
- Vérification ou Identification : c'est la phase de test ou bien la phase on-line, le même traitement est effectué dans les deux phases de manière à obtenir les vecteurs caractéristiques des images de tests qui sont calculés et projetés sur des espaces respectifs. Enfin, les scores de similarités sont calculés pour effectuer la classification et la décision finale.

Dans le système de reconnaissance de visage, les images des visages sont identifiées automatiquement en deux modes :

- **Vérification** : consiste en une correspondance un à un (1:1), qui compare une image du visage requête avec une image du visage modèle dont l'identité est proclamée. Ce système répondre à la question suivante : **suis-je ce que je prétends être ?**

- **Identification** : consiste en une correspondance un à plusieurs (1 : N) ou bien plusieurs à plusieurs (N : N), selon le protocole de la base de données (BDD). L'utilisateur ne présente pas une identité. Cependant, pose la question suivante : **qui suis-je ?**

Nous pouvons appeler les deux modes Identification ou Vérification : " *la reconnaissance* ". La figure 1.3 présente une vue d'ensemble d'un système de vérification et d'identification de visage. Durant la phase vérification, le code PIN (Personal Identification Number) d'un utilisateur est entré par l'intermédiaire d'un capteur biométrique qui capte la signature de l'individu et la convertit à une signature numérique compacte qui représente l'identité de l'utilisateur ou de l'individu. Cette signature est présentée à la fonction de classification pour faire la comparaison avec le modèle proclamé par l'utilisateur extrait de la base de données système à partir de son code PIN [21].

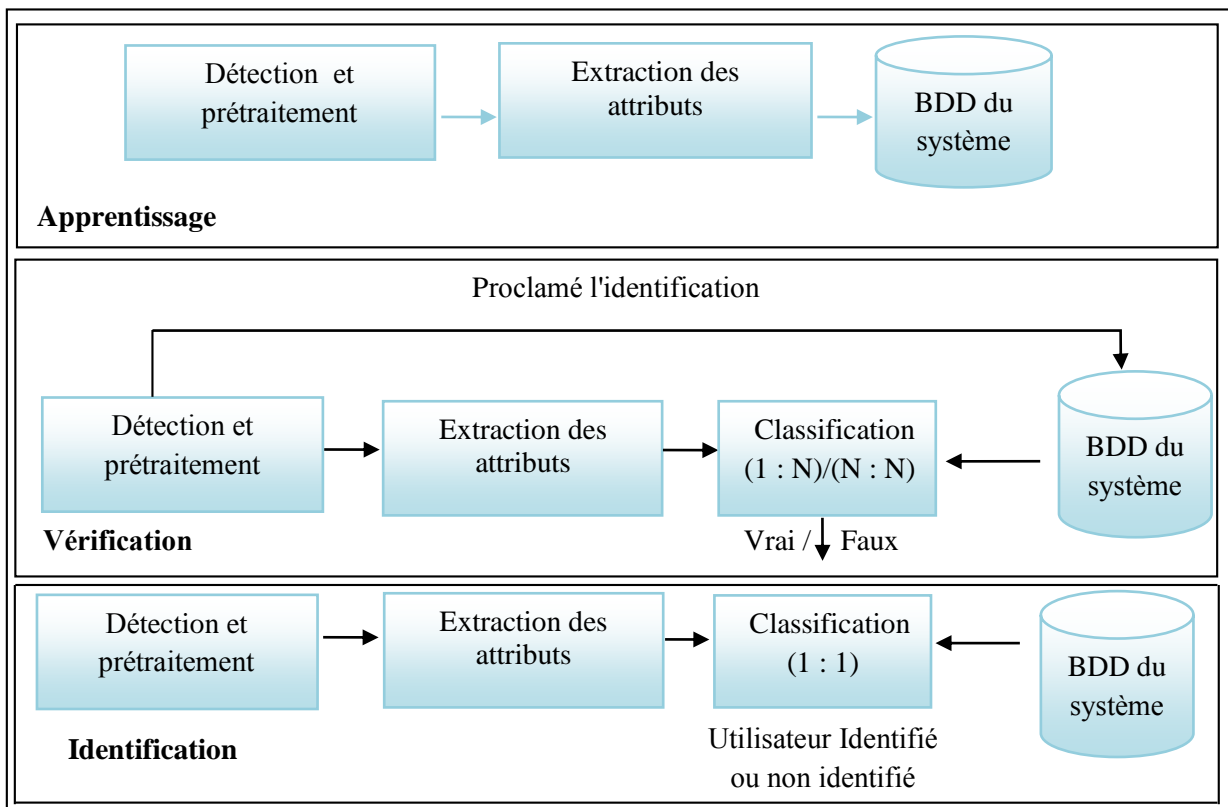


Figure 1.3 Schéma de fonctionnement d'un système biométrique.
Diagrammes des processus : apprentissage, vérification et identification

Dans la procédure d'identification, on n'a donc pas besoin d'un code PIN, le système d'identification compare la signature biométrique saisie avec les modèles de tous les utilisateurs dans la BDD système. La sortie est l'identité d'un utilisateur enrôlé ou un message, par exemple «utilisateur non Identifié» si l'individu n'existe pas dans la BDD système [21]. Les applications biométriques imposent une grande précision. Cependant, la procédure de reconnaissance on-line nécessite une exécution très rapide, parce que la majorité de ces applications requiert

également un temps de traitement court ; alors qu'il existe toujours un compromis entre le temps de calcul et la précision dans tous les systèmes biométriques.

1.2.3 Défis scientifiques de la reconnaissance automatique de visage

L'apparence de l'image de visage est affectée par un nombre de variations dans les conditions non contrôlées, cela, rend le traitement automatique de ces images dans un SRV très difficile. Un certain nombre de problèmes peuvent affecter les performances d'un SRV. Parmi ces difficultés, nous trouvons : la variation d'illumination, la variation à cause de l'expression faciale, la variation de poses, l'occlusion, la qualité d'image ...etc. Un résumé de ces défis est décrit ci-dessous.

1.2.3.1 La variation d'illuminations

Les conditions d'éclairage ambiantes ou l'illumination varient considérablement, notamment dans les environnements extérieurs. De fortes ombres peuvent être exprimées par une source de la lumière en raison de la structure 3D du visage humain. Les propriétés de la réluctance de la peau peuvent également aussi provoquer des variations d'illuminations dans les images de visage durant le processus de l'acquisition [22]. Les variations de la direction d'éclairage causent des variations drastiques dans l'apparence et la couleur de l'image faciale, ceci conduit à une faible extraction des caractéristiques faciales.

Il a été observé que des modifications provoquées par des variations dans l'illumination sont souvent plus grandes que les différences entre l'individu lui-même [23, 24] alors que le problème de discrimination entre les individus se pose fortement. Par conséquent, on trouve des difficultés dans le processus de la détection de visage et la classification au sein de la phase on-line du SRV 2 D. Ce qui rend la reconnaissance faciale basée sur les images 2D couleur une tâche très difficile, puisque l'information couleur et texture dépendent principalement des sources de la lumière. En revanche, la reconnaissance faciale 3D est plus robuste aux variations d'éclairage parce que l'information 3D est représentée par la géométrie faciale.

Buysens [25] en 2011 subdivise l'effet de l'illumination sur les images des visages en deux types : l'illumination globale ou ambiante et l'illumination locale. L'illumination globale affecte tout le visage de manière presque uniforme, d'autre part, l'illumination locale entraîne la création d'ombre et des zones éclairées [25]. Pour résoudre ce problème, un nombre d'approches qui traitent la présence des variations d'illumination dans les SRV est proposé dans la littérature. Pas mal de ces travaux sont basés sur l'extraction des caractéristiques faciales invariantes à la variation d'illumination ou bien le traitement de ce type de défi se fait par une

étape de prétraitement au début du SRV. La figure 1.4 présente un exemple du visage de différentes personnes (2 images par personne) avec des variations d'illuminations.

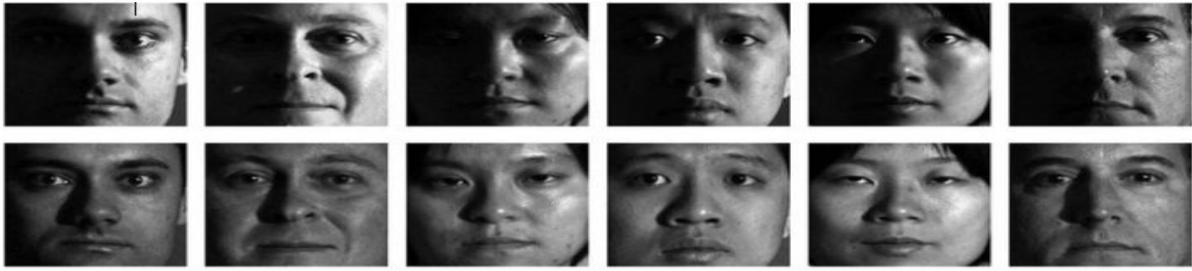


Figure 1.4 Exemple de variation d'illuminations pour différentes personnes

1.2.3.2 La variation d'expressions faciales

L'expression faciale est l'une des défis principaux pour les SRV. Le visage d'être humain est un objet non rigide, les différentes expressions faciales comme : le sourire, le rire, la colère, la surprise et la fermeture des yeux (Voir la figure 1.5) dégradent considérablement les performances de la reconnaissance de visage. Cela est dû au fait que l'expression faciale affecte l'information géométrique 3D et la couleur 2D du visage et en plus les positions des éléments faciaux tels que, la bouche les yeux et les joues. Cependant, le nez est la caractéristique faciale la plus stable et la plus invariante pour ces difficultés. Pour cela, plusieurs travaux de recherche [7, 26-29] se concentrent sur la zone nasale et notamment le bout du nez comme un point de référence pour détecter et reconnaître les personnes en présence des différentes expressions faciales.



Figure 1.5 Exemple de quelques variations d'expressions faciales de la même personne

1.2.3.3 La variation de poses

Dans de nombreuses applications de reconnaissance de visage, il y a toujours une différence dans l'angle de l'inclinaison de la tête entre l'image d'apprentissage et les images de tests, c.à.d. l'image d'apprentissage peut contenir une face frontale, tandis que l'image de test peut contenir un visage tourné avec un certain angle dans l'espace 3D (X, Y, Z). L'estimation de la pose de la tête est définie comme le calcul des trois types d'angles de rotation. Les trois degrés de liberté dans la pose de la tête sont nommé : angles Yaw ou angles de lacet c'est la rotation autour de l'axe Y (regarder à gauche ou à droite) ; les angles Pitch ou angles de la hauteur (regarder vers le haut ou vers le bas) c'est la rotation autour de l'axe des X et les angles

Roll ou angles de roulis, c'est la rotation autour de l'axe Z (la rotation à gauche ou à droite). La figure 1.6 présente les angles Yaw, Pitch et Roll pour la tête humaine.



Figure 1.6 Un exemple d'angles de rotation : Pitch, Yaw et Roll

L'une des grandes difficultés rencontrées par les techniques de reconnaissance faciale est la variation de poses [30]. Un grand nombre de ces systèmes se limitent aux poses frontales seulement ou des poses spécifiques requérant une estimation simple [25]. Tandis que, lorsqu'il y a de grands angles de rotation, l'estimation et la correction deviennent des processus nécessaires. Le domaine de recherche concernant l'analyse des mouvements de la tête et l'estimation de la pose est un domaine très actif. De nombreuses applications de l'interface de communication entre l'homme et la machine utilisent ce processus. Parmi ces applications [31]: les jeux vidéo, les techniques d'assistance et la commande d'un curseur d'ordinateur par des mouvements de tête, le domaine de la robotique (la commande des robots et des machines, la vidéosurveillance, la sécurité routière, l'analyse comportementale pour le marketing etc... La figure 1.7 illustre des exemples d'applications de l'estimation de pose de la tête.



Figure 1.7 Un exemple des applications pour l'estimation de la pose de la tête ; analyse d'émotion d'un conducteur de voiture ; la vidéosurveillance

L. Xiaoming et al [32] considère que la variation de poses est le type le plus difficile parmi les variations et les défis dans le SRV. Cela a été confirmé dans un autre travail proposé par *Gross et al* [33] qui ont montré que si le seul facteur de variation entre l'image enregistrée (apprentissage) et l'image-requête (test) est une rotation de la tête inférieure à 30° , les taux de reconnaissance des systèmes actuels sont de l'ordre de 90%. Des rotations plus importantes engendrent une forte baisse des performances [22] ; par contre une robustesse des SRV 3D au changement d'illumination. Ces performances sont également réduites en raison des données

manquantes causées par l'auto-occlusion. Par conséquent, la variation de la pose est le défi le plus difficile auquel sont confrontés les chercheurs récemment.

La figure 1.8 présente un exemple d'un visage subissant une rotation avec différents angles dans la base de données CASIA3DV4.



Figure 1.8 Exemples de variations de poses dans la base de données CASIA3DV4

1.2.3.4 Autres difficultés

Il existe d'autres types de difficultés qui peuvent apparaître pour un SRV. Parmi ces problèmes, on trouve par exemple, les occlusions figure 1.9 (a), la présence de maquillage, le vieillissement, les opérations chirurgicales sur la peau de visage et la présence ou l'absence des composantes structurelles (les moustaches, la barbe, les lunettes, le foulard), les images de visage à faible résolution ainsi que l'effet âge. La figure 1.9 (b) illustre le défi du vieillissement sur le visage d'Albert Einstein mais la constitution de bases de données s'étalant sur plusieurs années est difficile [25].



(a) Images en présence d'occlusions

(b) Images sur l'effet de l'âge

Figure 1.9 Exemples du vieillissement et de l'occlusion du visage

1.2.4 Principales bases de données

Afin d'évaluer la reconnaissance de visage, nous avons besoin d'une BDD dédiée à cette évaluation. Actuellement, il existe de nombreuses BDD de visage contenant les différents défis. Dans cette sous-section, nous introduisons les BDD de visage 2D et 3D les plus populaires. L'évaluation des différents algorithmes de reconnaissance de visage est toujours l'une des tâches les plus difficiles. Différents papiers peuvent utiliser que des parties de BDD pour leurs expériences. La partition des images d'apprentissage, les images de galerie et les images de tests peuvent également varier selon le protocole des différents scénarios de la reconnaissance. Le tableau 1.1 illustre les principales bases de données 2D et 3D.

Tableau 1.1 Les BDD 2D et 3D proéminentes et leurs caractéristiques

Nom de la BDD	Nbr des images	Nbr des Personnes	Type de données	Caractéristiques de la BDD			
				Ill	Exp	Pose	Autre
ORL	400	40	2D	Non	Oui	Oui	/
YALE	165	15	2D	Oui	Oui	Non	/
YALE B	5760	10	2D	Oui	Non	Oui	/
AR	3000	120	2D	Oui	Oui	Oui	longue période
CMU-PIE	41368	68	2D	Oui	Oui	Oui	/
FERET	10000	1000	2D	Oui	Oui	Oui	Longue période
FRVT	120000	30000	2D	Oui	Oui	Oui	Longue période
JAFFE	213	10	2D	Non	Oui	Non	/
XM2VTS	2360	295	2D	Oui	Non	Oui	/
LFW	13233	/	2D	Oui	Oui	Oui	Longe période
CASIA3DV4	4624	123	2D et 3D	Non	Oui	Non	/
Notre Dame	953	277	3D	Oui	Oui	Oui	Longe période
GavabDB	549	61	3D	Non	Oui	Oui	Occlusion
BU-3DFE	400	100	2D et 3D	Non	Oui	Non	/
BU-4DFE	60600	101	2D et 3D	Non	Oui	Oui	/
Bosphorus	4666	105	2D et 3D	/	Oui	/	Occlusion
FRGC v1	943	275	3D	Oui	Oui	Oui	/
FRGC v2	4007	466	2D et 3D	Non	Oui	Oui	Occlusion
FRAV3D	1696	106	3D	Non	Oui	Oui	/
MSU	533	90	3D	Non	Oui	Non	/
FSU	300	50	3D	Non	Oui	Non	/
ZJU-3DFED	360	40	2D et 3D	Non	Oui	Oui	/
University of York	5250	350	2D et 3D	Oui	Oui	Oui	/
IV ² multimodal	2880	365	2D et 3D	/	Oui	/	Qualité d'image
Texas	1149	118	2D et 3D	Non	/	Oui	/
3D_RMA	120	360	2D et 3D	Non	Oui	Oui	Occlusion

1.2.5 Évaluation et mesure de la performance

Il est particulièrement important de disposer d'une méthode d'évaluation de tels systèmes. Les systèmes biométriques doivent être évalués afin d'estimer les différents paramètres. La performance mesure l'efficacité et la fiabilité d'un système biométrique dans un contexte d'utilisation donné. Un ensemble des critères a été proposé dans la littérature pour mesurer les performances d'un système de reconnaissance de visage. Comme indiqué précédemment, il y'a deux scenarios de fonctionnement d'un système biométrique de reconnaissance de visage : identification et vérification, pour chaque mode il y'a différents critères pour évaluer le système. Dans cette section, nous allons introduire les taux, les erreurs et les courbes pour chaque mode.

1.2.5.1 Mesure des performances pour le mode vérification

Couramment, les critères de mesure de la performance utilisés pour le mode vérification sont :

- Le Taux de Faux Rejet (**TFR**) ou False Reject Rate (FRR) : c'est le pourcentage d'individus qui ont été faussement rejetés par le système alors qu'ils sont des clients.

$$TFR = (\text{Nombre de faux rejet})/(\text{Nombre de clients}) \quad (1.1)$$

- Le Taux de Fausse Acceptation (**TFA**) ou False Acceptation Rate (FAR) : c'est le pourcentage d'individus qui ont été acceptés par le système alors qu'ils sont des imposteurs.

$$TFA = (\text{Nombre de Fausse Acceptation})/(\text{Nombre d'imposteurs}) \quad (1.2)$$

- Le Taux d'Égale Erreur (**TEE**) ou Equal Error Rate (EER) est calculé pour vérifier la stabilité du système. Ce point de fonctionnement correspond à une valeur où FAR est égal à FRR. Le taux d'erreur TEE est un compromis entre le TFA et le TFR, c'est-à-dire, si un des deux taux est augmenté l'autre sera diminué automatiquement. En d'autres termes, plus l'EER est faible, plus le système est performant. À noter que ce taux d'erreur est le plus couramment utilisé dans la littérature pour illustrer la performance des systèmes biométriques [34].

- Le Taux de Vérification (**TV**) ou Vérification Rate (VR) : ce taux représente les individus qui sont correctement vérifiés par le système, soit les clients acceptés ou bien les imposteurs rejetés. Le taux de vérification permet d'examiner la sensibilité du système de vérification.

$$TV = (100 - (TFA + TFR)) \quad (1.3)$$

- La courbe **ROC** (Receiver Operating Characteristic): cette courbe constitue l'une des méthodes les plus couramment utilisées afin d'évaluer la performance globale d'un système d'authentification biométrique [34]. Il existe deux types de courbes ROC : celui qui trace le taux de vérification TV en fonction de TFA (Voir la figure 1.10 (a)) et celle qui trace le TFR en fonction de la TFA (Voir la figure 1.10 (b)).

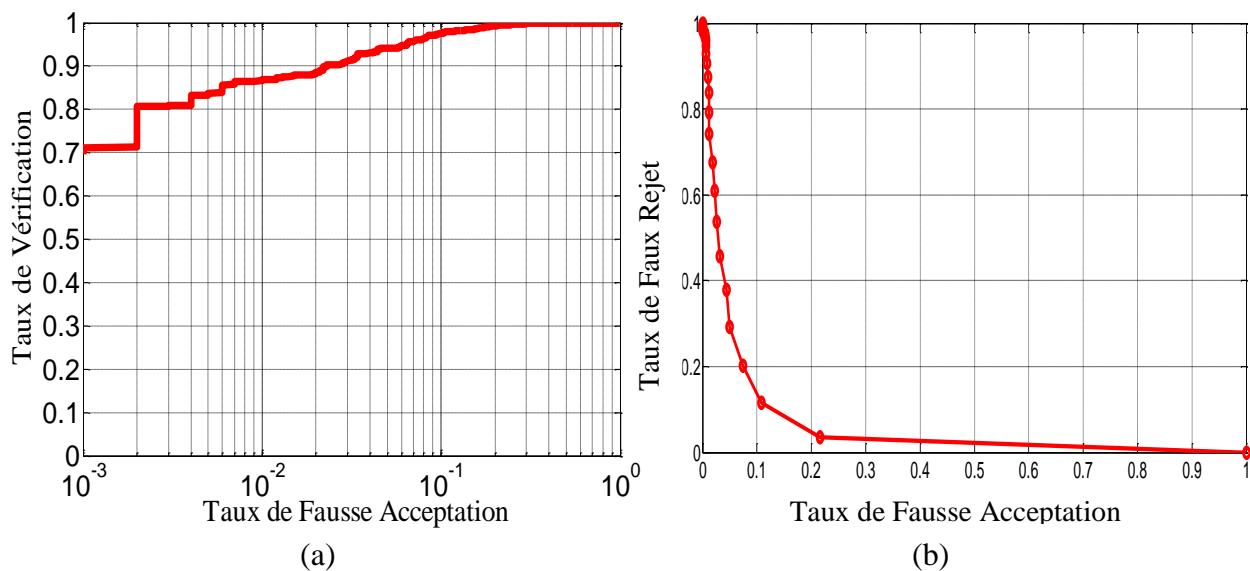


Figure 1.10 Exemples des courbes ROC ; (a) TV Vs TFA, (b) TFR Vs TFA

- Taux d'Erreur Moyenne aussi appelé **HTER** (Half Total Error Rate) : c'est la moyenne entre le TFA et le TFR.

1.2.5.2 Mesure des performances pour le mode identification

Couramment, les critères de mesure de la performance utilisés pour le mode identification sont :

- Rank-1 Recognition Rate (**R1RR**) : est le pourcentage de tous les individus pour lesquels le meilleur match dans la galerie appartient au même sujet. Le R1RR représente le pourcentage de toutes les entrées du système qui sont correctement identifiées.
- La courbe **CMC** (Cumulative Match Characteristic) : la courbe CMC trace le pourcentage d'individus reconnus correctement en fonction du nombre de rangs [22] qui est considéré comme une correspondance (match) correcte. Un exemple de la courbe CMC est illustré dans la figure 1.11.

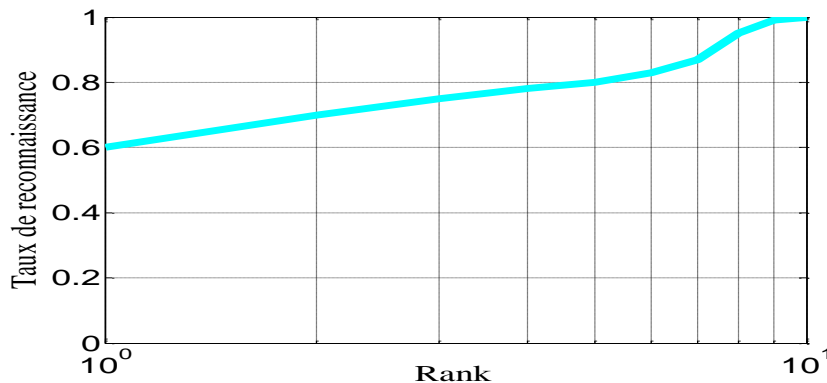


Figure 1.11 Exemple de la courbe CMC

1.3 Conclusion

Dans ce chapitre, nous avons décrit d'une façon générale le domaine de la biométrie et les différentes modalités utilisées pour reconnaître les personnes. Nous avons présenté le procédé du système de reconnaissance de visage, les différents problèmes et défis scientifiques rencontrés dans le monde réel, telles que la variation d'illuminations, d'expressions faciales et surtout la variation de poses de la tête comme un problème major. Nous avons introduit les bases de données de visage 2D et 3D les plus populaires et les plus utilisées dans la littérature pour l'évaluation des différents algorithmes de reconnaissance de visage. Nous terminons ce chapitre par une présentation des paramètres et des critères d'évaluation et de mesure de la performance pour les deux modes de reconnaissance de visage : vérification et identification. Dans le chapitre suivant, nous présentons un état de l'art des approches de la détection et la reconnaissance de visage 3D.

Chapitre 2

État de l'Art sur la détection et la RV3D

2.1 Introduction	21
2.2 De la reconnaissance de visage 2D vers le 3D	21
2.3 La détection de visage	22
2.3.1 La détection de visage en 2D	23
2.3.2 La détection de visage en 3D basée sur le bout du nez	24
2.4 Les techniques d'acquisition de données du visage 3D	28
2.4.1 Les techniques actives	28
2.4.1.1 Le scanner laser 3D	28
2.4.1.2 Le scanner 3D à lumière structurée	28
2.4.2 La technique passive	29
2.5 La représentation de visage 3D	29
2.5.1 La représentation par nuage de points	29
2.5.2 La représentation par le maillage 3D (3D Mesh)	30
2.5.3 La représentation par profondeur	30
2.6 Les différentes approches de la reconnaissance de visage 3D	31
2.6.1 Approches d'alignement des surfaces 3D	32
2.6.2 Approches basées sur le modèle 3D	33
2.6.3 Approches holistiques 3D	35
2.6.4 Approches Géométriques	36
2.6.5 Approches 2D+3D	38
2.7 Conclusion	40

2.1 Introduction

Un scan du visage 3D représente des informations faciales beaucoup plus discriminantes par rapport à la représentation 2D, car il fournit les informations réelles de la forme de visage humaine. Avec le développement rapide des équipements de capture 3D, il est devenu plus facile d'obtenir la représentation 3D. Cette modalité est considérée parmi les modalités les plus efficaces et les plus utilisées dans les systèmes de reconnaissance de visage.

La détection de visages dans l'image est un traitement indispensable et crucial avant la phase de reconnaissance. En effet, le processus de reconnaissance de visages ne pourra jamais devenir intégralement automatique s'il n'a pas été précédé par une étape de détection efficace. Le traitement consiste à rechercher dans une image la position des visages et de les extraire sous la forme d'un ensemble d'images ou zone visage dans le but de faciliter leur traitement ultérieur. La détection de visage peut se faire par détection de la couleur de la peau, la forme de la tête, détection du nez ou par des méthodes détectant les différentes caractéristiques du visage. Cette étape est d'autant plus délicate quand l'image acquise contient plusieurs objets de visage ou un fond non uniforme qui crée une texture perturbant la bonne segmentation du visage. Cette étape est dépendante de la qualité des images acquises.

Dans la première partie de ce chapitre, nous présentons une revue des approches de la détection et la reconnaissance de visage 3D. Nous étudions aussi les méthodes de la détection 3D basées sur le bout du nez comme un point de référence. Ensuite, nous présentons les techniques d'acquisition et les différents types de données pour représenter un visage humain sous forme tridimensionnelle. Enfin, nous présentons une classification des approches de la reconnaissance de visage 3D en cinq catégories.

2.2 De la reconnaissance de visage 2D vers le 3D

Récemment la reconnaissance du visage par la machine est devenue un sujet de recherche très important pour la communauté scientifique. Un grand investissement est fourni par les organismes gouvernementaux afin d'améliorer les systèmes de sécurité surtout ce qui concerne le contrôle des frontières et les aéroports. Les défis scientifiques de la reconnaissance de visage (voir le chapitre 1 section 1.2.3) ont été des sources de motivation pour introduire de nouveaux systèmes et techniques pour résoudre ces problèmes. La mise en œuvre d'un système automatique de reconnaissance de visage robuste et fiable est un verrou technologique qui n'est pas résolu à l'heure actuelle [35]. Plusieurs méthodes ont été proposées dans la littérature pour la reconnaissance de visage basée sur les images 2D. Cependant, les performances de ces

systèmes sont toujours affectées par un certain nombre de limitations dans les environnements non contrôlés. Ces dernières années, on parle de plus en plus de la reconnaissance de visage 3D comme une solution alternative pour résoudre les différents défis présents dans le monde réel [2]. La forme tridimensionnelle de visage est une modalité moins sensible à la majorité des problèmes et des défis cités précédemment, par conséquent plusieurs travaux de recherche se sont tournés vers l'étude de la modalité 3D [36]. L'acquisition de la forme 3D du visage est obtenue grâce à des dispositifs de numérisation 3D ou bien par des techniques de reconstruction [2]. Néanmoins, l'inconvénient le plus présenté des méthodes de reconnaissance de visage 3D est qu'elles traitent toujours le visage humain comme un objet rigide [37]; cela signifie que ces méthodes sont un peu sensible aux variations d'expressions faciales. Dans notre travail nous essayons de contribuer à surmonter ces défis par la proposition de quelques techniques efficaces pour un SRV3D et nD en milieux non contrôlés. Les différentes étapes pour la conception d'un tel système sont citées dans ce qui suit.

2. 3 La détection de visage

La détection des visages est une technique très importante pour la reconnaissance de visage et l'interaction homme-machine. Elle est la première étape pour tous les algorithmes d'analyse du visage, tel que : l'alignement, la modélisation de visage, la reconnaissance (vérification/identification), le suivi de la pose de la tête et le suivi de l'expression faciale [24]. Puisque les données faciales brutes peuvent contenir à la fois le visage humain et une grande zone d'arrière-plans, la détection de visage est utilisée pour segmenter précisément la région faciale d'intérêt. La détection des visages peut être effectuée en utilisant les images 2D, les données 3D ou bien une combinaison de celles-ci.

La détection de visage 3D est habituellement effectuée en se basant sur la localisation de bout du nez comme un point de référence [38]. De nombreux d'algorithmes de la détection de visage 2D sont disponibles dans la littérature [22]: les modèles de visage, le détecteur de visage de Viola et Jones [39], la détection de la peau et des différents algorithmes basés sur l'apprentissage [24]. Dans un autre classement, *Rodriguez, Yann et al* [40] en 2006 ont divisé les méthodes de détection des visages en deux catégories, la détection frontale de visage qui ne consiste pas une rotation de la tête (0°) et la détection non frontale de visage qui consiste des images faciales sous une rotation de la tête (Pitch, Yaw et Roll) avec différentes valeurs d'angles et différentes directions dans l'espace 3D.

2.3.1 La détection de visage en 2D

Les algorithmes de la détection de visage dans ce cas se basent sur des images 2D (images intensité couleur ou niveaux de gris) peuvent être classées en quatre groupes selon *Yang et al* [41] :

- **Approches basées sur les connaissances** (Knowledge-Based Methods) : ces méthodes se basent sur les caractéristiques du visage comme le nez, la bouche, les yeux et la relation entre eux pour classer les images visage et non visage. L'inconvénient de ce type de méthodes est qu'elles présentent un problème pour détecter le visage lorsque l'arrière-plan est complexe [35].
- **Les approches basées sur les caractéristiques invariantes** (Feature Invariant Approaches) : l'objectif de ces algorithmes est de trouver des caractéristiques structurelles qui existent même lorsque les conditions sont variées (l'éclairage, la pose, l'expression faciale etc.) tel que la texture ou la signature de couleur de la peau. Puis, ces caractéristiques invariantes ont été utilisées pour la localisation des visages en présence de ces problèmes.
- **Les approches de correspondance de modèle** (Template Matching) : l'idée de ces méthodes est de calculer la corrélation entre l'image d'entrée et un modèle pour la détection. Plusieurs modèles standards d'un visage sont stockés pour décrire le visage. Ces méthodes rencontrent encore quelques problèmes de robustesse liés aux variations de lumière et d'échelle [35].
- **Les approches basées sur l'apparence** (Appearance based methods) : ces méthodes utilisent le même principe des approches correspondance de modèle. Néanmoins, les modèles ici sont tirés d'un ensemble d'images d'apprentissage qui devrait capturer la variabilité que représente l'apparence du visage.

La méthode la plus utilisée dans la littérature pour la détection de visage en 2D est le détecteur Viola-Jones proposé en 2002 par *P. Viola et M. Jones* [42], ce papier est cité plus de 11750 fois selon Google Scholar. On peut dire que le détecteur Viola-Jones est une méthode qui appartient aux approches basées sur l'apparence dans la subdivision des méthodes de détection proposées par *Yang et al* [41]. Ce détecteur a récemment attiré une attention considérable en raison de sa haute performance, sa robustesse en temps réel et sa vitesse de traitement. L'algorithme de Viola-Jones consiste en trois étapes principales : l'intégrale de l'image, l'apprentissage basé sur l'algorithme Adaboost et la structure en cascade. Dans la première étape, l'image d'entrée est représentée sous forme d'intégrale pour extraire les caractéristiques de Haar afin d'augmenter la discrimination entre le visage et le non-visage. L'apprentissage basé sur l'algorithme Adaboost est appliqué pour former des modèles de

classification. Enfin, les modèles de classification sont regroupés dans la structure en cascade rapidement pour accélérer le processus de la détection. Cependant, de nombreuses travaux de recherche dans la littérature [43, 44] ont rapporté que la performance des caractéristiques d'Haar est encore très limitée dans les conditions difficiles en raison de sa simplicité. Récemment, d'autres variantes de caractéristiques ont été proposées telles que les caractéristiques de Haar conjointes [45] et les motifs binaires locales (LBP) [43]. Pour plus de détails sur le détecteur Viola-Jones voir [39]. Dans le cadre de cette thèse, nous nous intéressons en particulier aux techniques de la détection des visages 3D en utilisant les images de profondeur basées sur le bout du nez. La section suivante, présentera un état de l'art des techniques existantes, ce qui nous permettra de situer notre approche par rapport à ces travaux.

2.3.2 La détection de visage en 3D basée sur le bout du nez

Habituellement, toutes les approches de la reconnaissance des visages proposées dans la littérature tentent d'obtenir un bon taux de détection dans la première étape de leur système pour augmenter la précision et la robustesse. Dans les méthodes de reconnaissance faciale 3D, la région nasale joue un rôle très important dans la détection de visage [9, 46], la normalisation visage [38, 47], la correction de la pose [48, 49] et l'appariement basée sur la région du nez [50, 51]. Nous présentons ci-dessous une liste des références les plus pertinentes, nous analysons leurs techniques et discutons leurs limites.

Segundo et al [7] ont proposé une méthode pour la détection de bout du nez en présence de variation de poses basée sur les courbes faciales. Tout d'abord, ils trouvent la coordonnée y de bout du nez, suivie par les coordonnées x . Pour trouver les coordonnées y , ils calculent deux y -projections de profondeur, la courbe de profil et la courbe médiane, où la valeur de profondeur maximale représente la courbe de profil et la valeur de profondeur médiane représente la courbe médiane de chaque ensemble de points avec le même y de l'image de visage. Une autre courbe qui représente la différence entre le profil et les courbes médianes a été générée. Les auteurs supposent que le bout du nez est un pic à la courbe de profil à l'intérieur d'un tronç qui a la somme maximale de valeurs de différence (la courbe de différence). Ensuite, pour trouver les coordonnées x du point de bout du nez, la projection x de l'image de courbure est calculée. Ceci est fait par le pourcentage des points de crêtes de courbure de 7 lignes voisines centrées dans le nez. Un taux de détection de 99,95% a été obtenu. La figure 2.1 représente quelques exemples de la détection de bout du nez avec différentes poses. Cependant, le point faible de cette méthode est qu'il présente une faiblesse dans les angles de rotation selon l'axe Y et Z . *Anuar L.H et al* [29] ont proposé une approche pour détecter la région de bout du nez pour un modèle

de visage 3D en utilisant la courbure de la géométrie locale et le point signature. Basé sur des points convexes de la surface de visage, les auteurs utilisent une approche morphologique pour obtenir les candidats de la région de bout du nez constituant les points de plus haute densité. Une signature est générée et évaluée avec l'apprentissage de la bande de tolérance de bout du nez à des fins de correspondance. La région des points avec le meilleur score est choisie comme une région finale de bout du nez. Un taux de détection égal à 95.19% a été obtenu parmi plus de 1300 modèles de visage 3D. L'approche proposée est schématisée à la figure 2.2

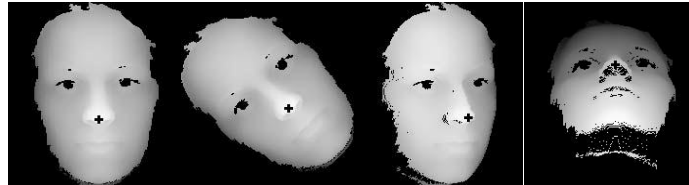


Figure 2.1 Exemple de la détection de bout du nez avec différentes poses [7]

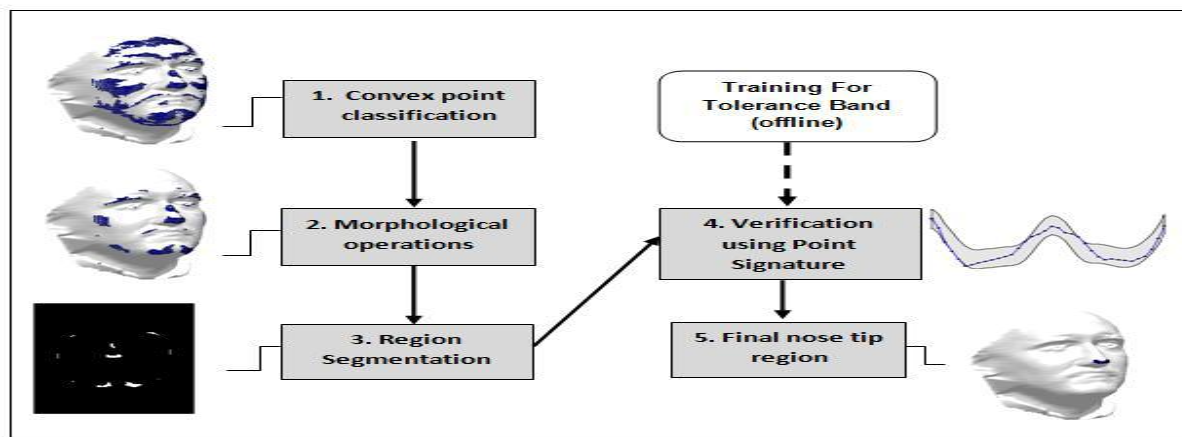


Figure 2.2 L'approche proposée par Anuar L.H et al pour la détection de bout du nez [29]

Chenghua Xu et al [28] ont présenté un système de filtrage hiérarchique comprenant trois phases (comme illustré dans la figure 2.3) basé sur la combinaison des caractéristiques locales et les caractéristiques statistiques pour détecter les positions de bout du nez. Ensuite, la crête du nez est estimée par une courbe appelée IAC (Included Angle Curve). La notion d'énergie efficace est introduite pour décrire la distribution locale des points voisins de bout du nez. Le SVM est utilisé pour sélectionner le point correct de bout du nez. Les résultats expérimentaux montrent que le taux de détection obtenu est de 99,3%. L'inconvénient de cette méthode est qu'elle nécessite un apprentissage pour le classificateur SVM donc un temps de calcul supplémentaire pour le système de reconnaissance. Par contre, la méthode proposée par Peng et al dans [26] est une méthode efficace pour la détection de bout du nez basée sur l'information de profondeur 3D sans l'exigence de l'apprentissage. Tout d'abord, les profils le plus à droite (-90°) et le plus à gauche ($+90^\circ$) du visage sont obtenus par la rotation de l'image de profondeur

sur un intervalle de $[+90^\circ, -90^\circ]$ avec un pas de 3° . La prochaine étape est de sélectionner les possibles candidats de bout du nez, dans laquelle le centre d'un cercle de rayon r est déplacé le long du profil de visage. Finalement, l'identification de bout du nez est effectuée par le calcul de deux mesures, le fitness du point cardinal et le fitness de pic. Un taux élevé de détection de 99,43% a été obtenue.

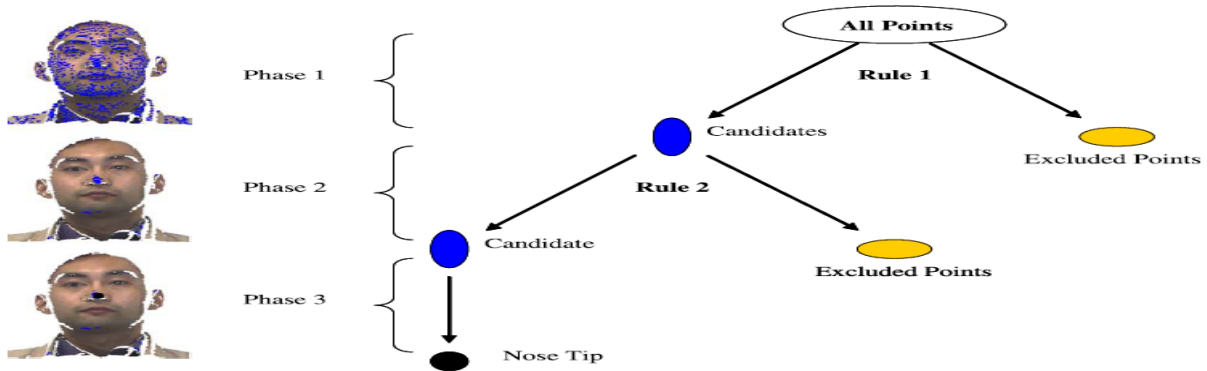


Figure 2.3 Le système de filtrage hiérarchique pour la détection de nez [28]

Mian et al [38] ont proposé une méthode de détection de bout du nez basée sur le découpage de l'image de visage en tranches horizontale et verticale. Pour chaque contour de tranche, un triangle d'altitude maximale est trouvé en utilisant un cercle dont le centre se déplace le long du contour de la tranche. Les sommets du triangle sont tous sur le contour de la tranche avec un sommet qui coïncide avec le centre du cercle et les deux autres sommets étant les deux sections du cercle entre le contour de tranche (voir la figure 2.4). Les candidats de bout du nez pour ce contour de tranche sont considérés comme le centre du cercle. L'échantillonnage aléatoire consensus (RANSAC) est utilisé pour éliminer encore les valeurs aberrantes. Celui qui a le maximum de confiance est considéré comme le bout du nez. Lorsque cette méthode repose sur l'altitude de triangle pour déterminer le bout du nez, il est susceptible à certaines variations de poses. La précision de l'algorithme proposé est de 98,3%.

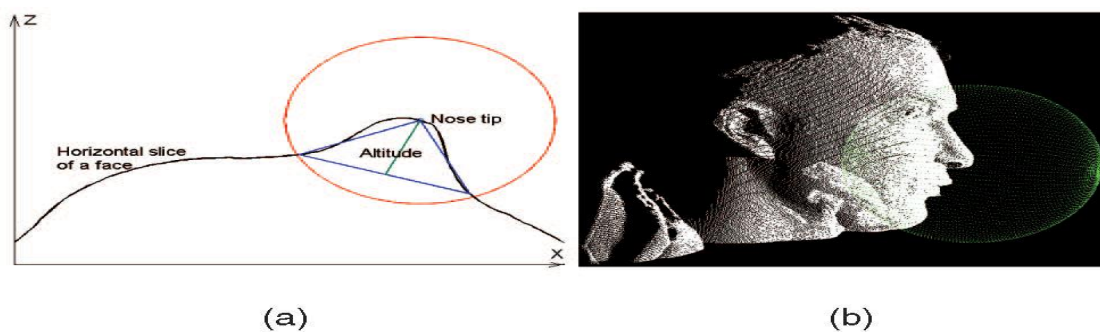


Figure 2.4 Illustration de la méthode de *Mian*,

- (a) la détection de la pointe du nez,
 (b) sphère centrée sur le bout du nez d'un nuage de points 3D pour recadrer le visage [38]

Rong Liu et al [52] dans leur papier, proposent une approche de détection de nez basée sur le Template Matching d'images de profondeur. Pour cela, ils peuvent détecter la région du nez même dans des situations difficile telles que les occlusions et l'absence d'une partie du visage dans lequel la propriété symétrique du visage est détruite. Avec l'utilisation de la méthode de corrélation normalisée croisée (CCN), la procédure Template Matching a un temps de traitement efficace, qui permet une recherche rapide des directions frontales et une robustesse contre les expressions et les variations de poses.

Une nouvelle méthode de détection de bout du nez est proposée par *Ye Li, et al [46]* basée sur la fusion des caractéristiques de courbure d'ordre bas et d'ordre supérieur. Tout d'abord, les points de crête sont détectés à partir des surfaces locales 3D de visage par le calcul des caractéristiques de courbure supérieure, à savoir les dérivées des courbures principales le long de leurs directions de courbures correspondantes. Ensuite, des lignes de crête de candidats de nez sont choisies à partir de toutes les lignes des crêtes convexes à la fois, selon le nombre de points de crête sur chaque ligne et selon l'intensité de crête de chaque ligne. Enfin, le bout du nez se trouve en fonction des caractéristiques de la distribution de courbure gaussienne autour des points de crête (sur tous les candidats de la ligne de crête du nez). L'inconvénient de cette méthode est qu'elle n'est pas purement automatique, les auteurs utilisent un découpage manuel pour la zone de visage, on peut dire que cette méthode n'est pas efficace sur la base de données utilisée pour l'évaluation (GavabDB). *Slater et al [53]* ont utilisé un vecteur de caractéristiques contenant une intégrale des variantes de multiple échelle, puis l'analyse discriminante quadratique pour classer les vecteurs de caractéristiques de bout du nez et non-nez.

Alessandro Colombo et al [54] ont proposé une méthode basée sur la détection des caractéristiques saillantes du visage tels que les yeux et le nez par une analyse de la courbure de surface. Les auteurs dans ce papier recherchent la région du nez dans les régions positives dans la carte de la courbure moyenne seuillée et ne tenant compte que des régions concaves elliptiques. Basé sur un triangle de surface formé par un candidat de nez et deux candidats des yeux, où chaque triplet (triangle) est traité par un classificateur PCA formé pour la discrimination entre les visages et les non-visages. Cette méthode est très sensible pour les rotations selon les axes X et Y. Une autre méthode similaire proposée par *K.I.Chang et al [50]*, les auteurs créent une région de segmentation basée sur le type de courbure, puis ils détectent des repères sur le visage. Ils ont mis deux seuils, l'un pour voir la carte de courbure gaussienne et l'autre pour la carte de courbure moyenne. Enfin, le bout du nez est trouvé comme un pic.

Cependant, la méthode se base sur les yeux et le nez comme repère de détection ce qui est toujours faible en présence de variations de poses.

2.4 Les techniques d'acquisition de données du visage 3D

Récemment, différentes techniques d'acquisition des données 3D sont proposées par les chercheurs dans ce domaine. Ces techniques sont intensivement utilisées dans : la biométrie faciale, la médecine, la réalité virtuelle, la robotique et la reconnaissance de formes. Les techniques d'acquisition des données 3D peuvent être classées en deux grandes familles, les techniques actives utilisant un scanner 3D et les techniques passives qui sont tentées de déterminer la forme de la scène par reconstruction de modèle 3D à partir d'une ou plusieurs images 2D.

2.4.1 Les techniques actives

Les techniques actives sont les techniques les plus utilisées pour acquérir la forme 3D de visage [2]. Les dispositifs actifs utilisent un capteur photographique avec une source lumineuse spécifique permettant d'illuminer l'objet, afin de mesurer les coordonnées tridimensionnelles des points sur la surface de cet objet. Les techniques actives les plus connus [2, 8, 36] sont brièvement décrits ci-dessous.

2.4.1.1 Le scanner laser 3D

Ce scanner est basé sur le processus de la triangulation et donne l'information sur la profondeur de l'objet [8]. Parmi ces scanners les plus populaires : le Minolta Vivid 910 de Konica Minolta (voir figure 2.5 (a)), et le Cyberware 3030 (voir figure 2.5 (b)). L'une des limitations principales de la technique de triangulation est qu'il exige un temps relativement long pour acquérir les données 3D et durant cette période d'acquisition la personne doit rester immobile [55].

2.4.1.2 Le scanner 3D à lumière structurée

Le principe de ces techniques est la projection d'une lumière monochromatique structurant sous forme d'une bande, grille ou des motifs elliptiques. La géométrie de la scène est obtenue en analysant les déformations du motif lumineux ou en décodant la succession de la lumière envoyée sur l'objet à numériser. La technique de la lumière structurée est capable d'acquérir des séquences d'images 2D et 3D en temps réel. Néanmoins, les données 3D acquises viennent avec un certain nombre de trous et des défauts causés par la réflexion élevée ou une réflexion

faible en raison des surfaces sous-jacentes [22]. Le Microsoft Kinect 3D est un exemple des scanner à lumière structurée (voir figure 2.5 (c)).

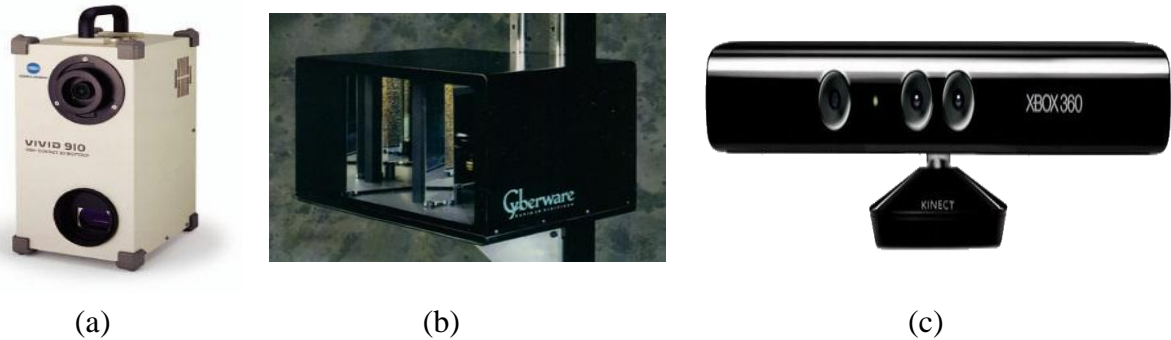


Figure 2.5 Les scanners 3D les plus populaires ;
(a) Minolta Vivid 910, (b) Cyberware 3030, (c) Microsoft Kinect

2.4.2 La technique passive

Les techniques de la reconstruction 3D ou les techniques passives utilisent la reconstruction 3D de la scène à partir d'une ou plusieurs images avec deux caméras 2D calibrées. Plusieurs méthodes sont proposées pour la reconstruction 3D : les méthodes basées sur le modèle 3D, les méthodes basées sur le mouvement (Shape From Motion, SFM), les méthodes basées sur les ombres portées (Shape From Shading, SFS) et les méthodes de stéréovision [2]. L'inconvénient de ces méthodes est qu'elles consistent des calculs complexes puisque le modèle de visage 3D est construit à partir de plusieurs images 2D. Pour plus de détail concernant les différentes méthodes et techniques de l'acquisition 3D voir [2].

2.5 La représentation de visage 3D

La représentation du visage 3D est très importante pour le choix de l'approche et les méthodes de la reconnaissance utilisées. Dans cette section, nous décrivons les méthodologies des différents types de représentation faciale 3D en mettant en évidence la représentation par l'image de profondeur celle utilisée dans notre contexte de travail.

2.5.1 La représentation par nuage de points

Le nuage de points est une collection de l'ensemble des coordonnées 3D (X, Y, Z) des points du visage non ordonnés. Le visage humain est représenté par N échantillons sous forme de trois vecteurs de coordonnées X, Y et Z de longueur N. La représentation de nuages de points est la forme la plus simple pour représenter la surface faciale. C'est également la représentation la plus commune générée par les scanners 3D [56]. La figure 2.6 montre la représentation par

nuage de points pour un visage que nous avons obtenu à partir de la base de données CASIA3DV4.

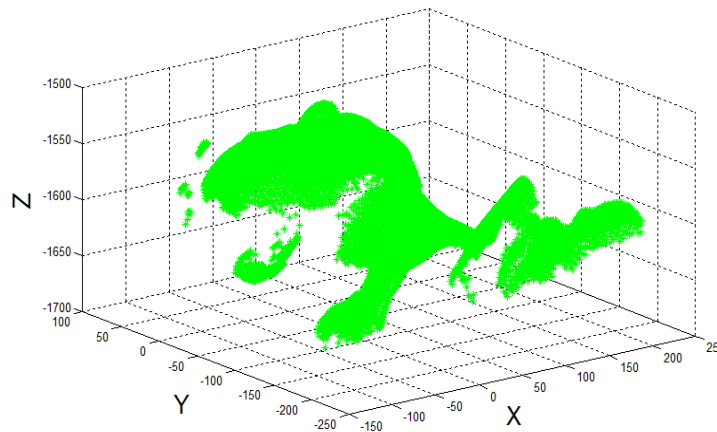


Figure 2.6 Représentation par nuage de points pour un visage CASIA3DV4

2.5.2 La représentation par le maillage 3D (3D Mesh)

Un maillage 3D est un ensemble de sommets et d'arêtes qui représente la forme d'un visage dans l'espace. Le maillage de visages se compose généralement, de quadrilatères, des polygones convexes simples ou des triangles. Ce dernier est le type de maillage le plus utilisé [22]. Grâce aux informations délivrées par la représentation maillage 3D, il est possible de mesurer les distances géodésiques entre les points (sommets et d'arêtes) du visage en utilisant les informations de connectivité. Une représentation de visage avec le maillage 3D basée sur la triangulation est illustrée dans la figure 2.7.

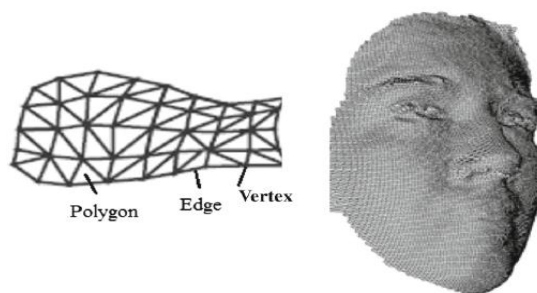


Figure 2.7 Représentation par maillage 3D pour un visage GavabDB [8]

2.5.3 La représentation par profondeur

Une autre méthode populaire pour représenter la forme du visage consiste à projeter les données de profondeur du visage en image niveaux de gris 2D. L'image de profondeur également appelée l'image de gamme ou bien l'image 2.5D correspond à une représentation bidimensionnelle d'un ensemble de points 3D (X, Y, Z), où chaque pixel dans le plan (X, Y) stocke la valeur de sa profondeur z. Les valeurs sont généralement normalisées pour donner une image en niveaux de gris où les pixels noirs correspondent au fond tandis que les pixels

blancs représentent les points de surface les plus proches à la caméra ou du scanner. Le système de coordonnées d'un visage 3D et son image de profondeur dans le plan (X, Y) sont présentées dans la figure 2.8.

Les images de profondeur constituent une bonne alternative au maillage 3D. Cette représentation est utilisée par nombreux scanners 3D. Ce type de données présente l'avantage d'applicabilité dans les méthodes de reconnaissance de visage 2D grâce à la projection de la forme tridimensionnel du visage sur une image de niveaux de gris. Récemment, plusieurs systèmes de reconnaissance de visage basés sur les images de profondeur 3D ont été proposés [10, 11, 43, 57-59] et ont obtenus une grande précision dans des conditions non contrôlées tels que la variation d'illumination, d'expression faciale, d'occlusion et de variations de pose. Dans notre travail, nous utilisons la représentation du visage humain par la profondeur pour développer un système de reconnaissance de visage robuste et efficace en présence d'illuminations, d'expressions et surtout une grande variation de poses.

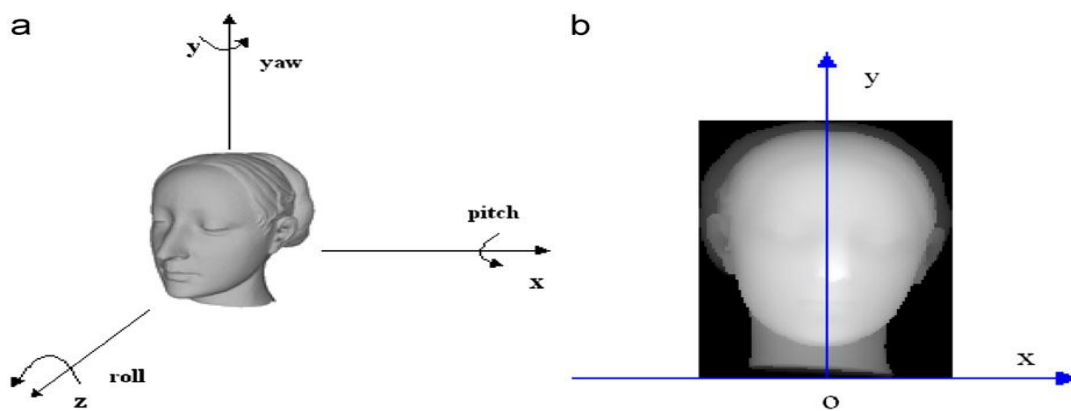


Figure 2.8 Représentation de visage par profondeur ;
(a) système de coordonnées 3D, (b) image de profondeur dans le plan (X,Y) [39]

2.6 Les différentes approches de la reconnaissance de visage 3D

La reconnaissance 3D de visage constitue une alternative prometteuse aux différents facteurs de variabilité qui se posent dans la reconnaissance faciale 2D. Récemment, les travaux de recherche se basent sur l'analyse des données 3D qui conservent toutes les informations sur la géométrie du visage, ce qui permet d'avoir une représentation réelle de ce dernier. Les approches 3D de reconnaissance du visage peuvent être divisées en cinq catégories :

- (1) *Approches d'alignement des surfaces 3D*,
- (2) *Approches basées sur le modèle 3D*,
- (3) *Approches holistiques 3D*,
- (4) *Approches Géométriques*,
- (5) *Approches 2D+3D*.

2.6.1 Approches d'alignement des surfaces 3D

Le problème pour ces méthodes est l'alignement de deux surfaces de visage 3D afin de faire la comparaison entre elles (voir la figure 2.9). Nombreux systèmes de reconnaissance faciale 3D introduits jusqu'à présent, utilisent l'algorithme ICP (Iterative Closest Point) ou la plus proche voisine itérée qui a été introduite par *Besl et McKay* [60] en 1992. Ce papier est cité plus de 11726 fois selon Google Scholar, cela montre le rôle et l'importance de cet algorithme. L'ICP peut être utilisé non seulement dans le domaine de la reconnaissance de visage, mais aussi dans de différents domaines tels que la simulation de la chirurgie et la conception assistée par ordinateur (CAD) pour les objets de la modélisation [61].



Figure 2.9 Alignement de deux visages 3D [2]

L'algorithme ICP transforme un nuage de points 3D "B" pour l'alignement avec un autre nuage de points 3D "A" (modèle de référence). Une série de transformations incrémentales est effectuée à plusieurs étapes de façon itérative. Au cours de chaque étape d'itération, l'erreur quadratique moyenne (MSE) entre deux nuages de points est calculée. L'algorithme est convergent lorsque la valeur de la MSE est minimale. L'inconvénient de cet algorithme est le besoin d'un modèle de référence comme une condition initiale, aussi le temps de calcul est très élevé pour la comparaison entre le visage de test et tous les visages d'apprentissage de la base de données. D'autre part, les méthodes basées ICP sont sensibles au problème du changement d'expressions, car elles considèrent le visage 3D comme un objet rigide [2, 36]. Plusieurs méthodes qui utilisent l'algorithme ICP et ses variantes pour l'enregistrement de visage sont présentés dans la littérature. Une approche basée sur ICP pour aligner deux surfaces de visage est donnée par *Meduoini et Waunpot* [62], cette approche est en contraste avec les méthodes de reconnaissance du visage traditionnel, qui comparent les images de visages afin de reconnaître les personnes. Les auteurs ont également analysé les exigences de la qualité de l'image afin de générer une bonne qualité de reconstruction 3D. Ils ont également mené des expériences sur les images de 100 individus, avec 7 poses différentes pour chaque individu, un TEE inférieur à 2% est obtenu.

Ben amor et al [63] ont proposé une approche pour l'appariement (la correspondance) des visages en utilisant un modèle partiel 2.5D avec des expressions faciales et différentes variations de l'angle de la tête. Premièrement, une approximation des transformations entre les points de vue (poses) est utilisée, les auteurs appliquent un alignement grossier, puis ils effectuent un alignement par ICP entre les points dans une vue et les points les plus proches, les résultats sont deux ensembles de points dans le modèle 2.5D de test et le modèle de visage 3D à partir de la base de données. La déviation spatiale globale entre chaque paire de points est fournie. Le processus de reconnaissance est basé sur la distribution obtenue de cette distance produite par région.

Pour résoudre le problème de la sensibilité aux variations d'expressions faciales *Moreno et al* [64] ont proposé un système automatique de reconnaissance 3D de visage en présence d'illuminations, de poses et d'expressions faciales en utilisant la segmentation de visage en plusieurs régions basée sur les signes de la courbure médiane et la courbure gaussienne afin d'isoler les régions de courbures prononcées. Ensuite, un vecteur caractéristique est créé pour chaque personne. D'autre part, pour résoudre le problème du temps de calcul, *Bardsley et al* [65] ont proposé une approche de reconstruction et de reconnaissance 3D de visage. Ils ont créé un seul modèle de référence commun par le calcul de la moyenne de tous les visages de collection. Ensuite, le visage de test est aligné avec ce modèle de moyenne. Une précision de 98,2% a été obtenue en utilisant une base de données contenant plus de 175 mailles.

2.6.2 Approches basées sur le modèle 3D

L'idée principale de ces méthodes est de construire un modèle tridimensionnelle de visage humain avec ces différentes variations. Cette construction est réalisée grâce à une localisation de certains traits du visage telle que les coins ou les centres des yeux, les sentiers de nez, le bout du nez etc... *Blanz et al* [66, 67] ont proposé un système de reconnaissance de visage basé sur un modèle bien connu dans la littérature nommé 3D Morphable Model (3DMM). Cette approche est déjà utilisée dans divers systèmes de reconnaissance de visage commerciaux [68]. Le 3DMM est utilisé pour capturer une véritable géométrie tridimensionnelle de la surface faciale qui est utilisée en combinaison avec les informations de l'apparence du visage afin d'effectuer les comparaisons, ce qui paraît être une meilleure représentation des visages humains. Le 3DMM utilise plusieurs points de repère pour initialiser les paramètres et ensuite les optimiser par une optimisation non linéaire (voir figure 2.10). Ce modèle est le modèle le plus populaire pour l'estimation de la pose, l'éclairage, la forme et les paramètres de texture de l'image faciale [69]. Cependant, pour une image de visage, le 3DMM généralement requiert un

temps de calcul relativement long (plusieurs minutes pour obtenir un bon résultat). Pour traiter ce problème, *Dong Yi et al* [69] ont présenté un algorithme rapide appelé le filtre adaptatif des poses ou "Pose Adaptive Filter" (PAF) basé sur les paramètres de la pose et de la forme (voir figure 2.11), et les autres paramètres sont négligés. Par rapport à l'algorithme 3DMM, PAF est moins précis, mais il est assez bon pour la reconnaissance faciale en présence de poses.

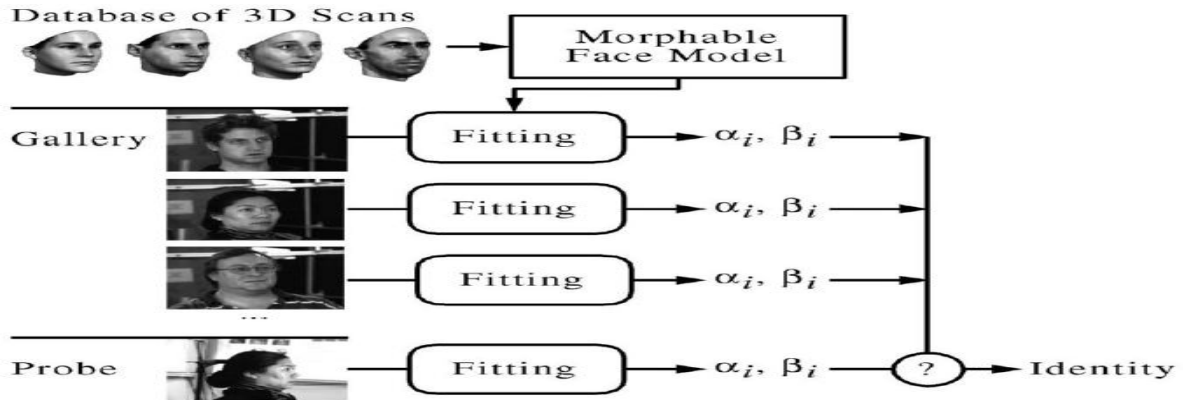


Figure 2.10 Le modèle 3DMM proposé par *Blenz et al* [67]

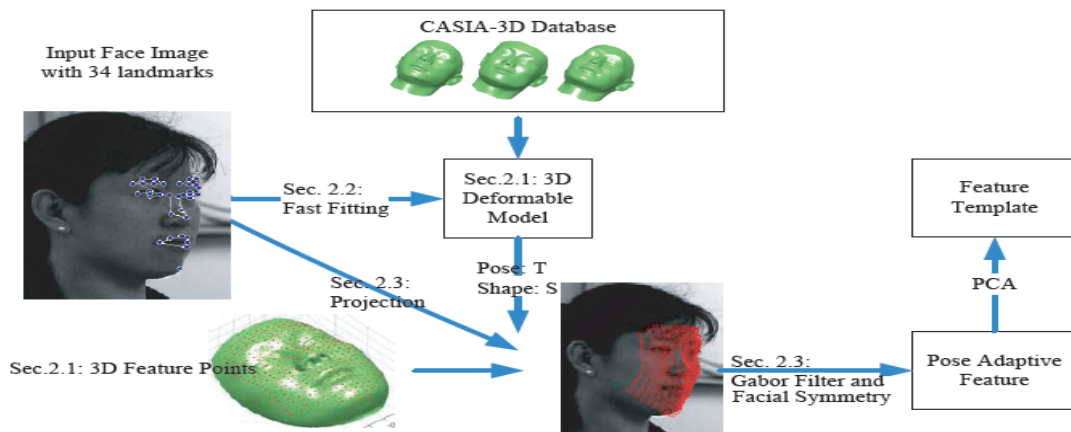


Figure 2.11 L'organigramme du modèle PAF proposé par [69]

Dans le même contexte, *Zhou et al* [70] présentent une modèle 3D de forme active ou "3D Active Shape Model" (3DASM) pour localiser automatiquement 58 points de caractéristiques (Landmark) de visage à partir de différents points poses. Le 3DASM est formé sur des données symétriques, générées en utilisant un 3DMM. Puisque les données d'apprentissage sont utilisées directement à partir de 3DMM, les Landmark ont une correspondance une à une entre les points 2D détectés à partir de l'image faciale et les points 3D sur le 3DMM. Le point faible de ces méthodes est la complexité de la construction d'un modèle 3D déformable. Elles nécessitent la collection d'un grand nombre de scans 3D de visage ainsi que les annotations d'une façon manuelle [2, 71].

2.6.3 Approches holistiques 3D

Les méthodes holistiques sont des méthodes qui utilisent le visage entier sous forme d'une image de profondeur, une maille ou un nuage de points comme une entrée pour le système de reconnaissance. Ces méthodes sont appelées aussi les méthodes globales ou les méthodes basées sur la réduction de la dimensionnalité. Des techniques statistiques telles que PCA et LDA et ICA sont largement utilisées pour la reconnaissance de visage basée sur les images faciales 2D. Plus récemment, des approches basées sur l'algorithme PCA ont été également appliquées aux données de visage 3D [9, 72-77] pour générer les vecteurs caractéristiques qui sont utilisés pour la reconnaissance. L'algorithme PCA ou la méthode Eigenfaces est considérée comme une référence de base pour l'évaluation des algorithmes de reconnaissance de visage 3D. La PCA peut être utilisée pour traiter les expressions faciales en incluant des visages expressifs dans l'ensemble de données d'apprentissage [78]. *Al-Osaimi et al* [77] ont produit des modèles de déformations d'expressions faciales en utilisant les vecteurs propres à partir des données d'apprentissage basées sur la représentation tensorielle. Ces modèles appris sont ensuite utilisés pour transformer les déformations des expressions. Par conséquent, les déformations faciales causées par les expressions sont séparées par les disparités interpersonnelles, rendant le système de reconnaissance de visage plus robuste contre les variations d'expressions faciales. Le nuage de points est représenté par un tenseur multiple de rang 0. Chaque trame locale de tenseur est intégrée à tout le domaine global dans un histogramme 2D qui est indexé par un champ local dans une dimension et un champ global dans l'autre dimension. Enfin, les coefficients PCA des histogrammes 2D sont concaténés dans un vecteur caractéristique unique qui représente la signature biométrique de la personne (voir figure 2.12).

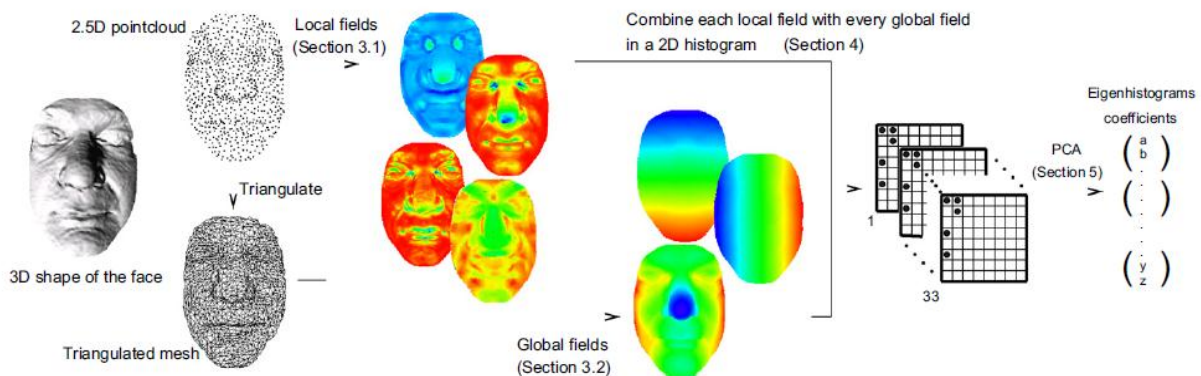


Figure 2.12 Illustration de l'approche proposée par *Al-Osaimi et al* [77]

L'algorithme LDA est un algorithme d'apprentissage supervisé très utilisé dans la littérature. L'algorithme Fisherfaces 3D produit des résultats significativement meilleurs que

l'algorithme Eigenfaces 3D [79]. L'outil Fishersurfaces 3D [80] est une extension de la méthode Fisherfaces basée sur l'algorithme LDA 3D appliqué aux données faciales 3D maillées. La LDA peut également être utilisée en combinaison avec la PCA ou autres méthodes d'extraction, comme décrit dans [81, 82]. *Gupta et al* [83] ont proposé un système de reconnaissance de visage 3D basé sur le calcul des distances géodésiques et euclidiennes, les auteurs utilisent 600 distances (300 distances géodésiques et 300 distances euclidiennes) entre 25 paires de points de référence localisées manuellement comme des caractéristiques. Ensuite, LDA est utilisée pour réduire les dimensions de ces caractéristiques.

L'Analyse en Composantes Indépendantes (ICA) [72] est une généralisation de l'algorithme PCA qui a été largement utilisé pour la projection de sous-espace [84, 85]. Il projette les données faciales 3D sur un ensemble de vecteurs de base statistiquement indépendants. L'ICA a été utilisée pour la reconnaissance de visage 3D à partir d'images de profondeur. En plus des moments statistiques du deuxième ordre (le cas de PCA), ICA utilise aussi les moments statistiques d'ordre supérieur. Par conséquent, L'ICA donne une représentation plus discriminante que PCA. Une grande quantité d'informations importante d'une image est codée sous forme de relations statistiques d'ordre supérieur [22, 86]. *Zhaoxiang et al* [87] ont développé une nouvelle méthode pour la reconnaissance 3D du visage basée sur les caractéristiques harmoniques sphériques (Spherical Harmonic Features, SHF). Les visages 3D sont caractérisés par les énergies contenues dans la carte de profondeur sphérique (SDM), et une représentation de SHF est ensuite générée à partir de la SDM avec des fréquences différentes. La forme brute des surfaces faciales 3D est codée par les énergies aux basses fréquences, tandis que les détails fins sont capturés par les énergies aux hautes fréquences. Cette méthode présente une légère baisse des performances lorsque l'expression faciale et les occlusions sont prises en compte.

2.6.4 Approches Géométriques

Ces méthodes sont basées sur les caractéristiques faciales locales géométriques ainsi que les différentes courbes faciales. L'utilisation de la géométrie faciale 3D pour les systèmes de reconnaissance de visage est considérée comme une voie très prometteuse. Néanmoins, la variation d'expressions faciales est un défi encore ouvert pour ces approches [36]. Une méthode globale basée sur la courbure a été présentée par *Wang et al* [88], la surface d'un modèle du visage a été représentée par une image gaussienne étendue (Extended Gaussian Image, EGI) afin de réduire le problème de la reconnaissance de visage 3D à une comparaison de l'histogramme 2D. La première étape pour extraire la caractéristique EGI est de définir une

sphère gaussienne générée par la division d'une sphère en un certain nombre de cellules. L'EGI représente les régions convexes qui sont insensibles pour la variation d'expressions faciales. Un autre type de méthodes géométriques est basé sur les profils de visage 3D. *Beumier et Acheroy* [89] ont proposé d'utiliser le profil vertical des modèles 3D pour la reconnaissance du visage. Leur première étude a été basée sur trois profils d'un visage et a eu un taux d'erreur de 9,0%, un ensemble de données de 30 sujets ont été utilisés pour le test. Les mêmes auteurs ont ajouté des informations de niveaux de gris au processus d'appariement [90], cette tentative a réduit le taux d'erreur à 2,5% quand la méthode a été testée sur la même base de données. *Drira et al* [91] représentent les surfaces nasales du visage 3D en utilisant des collections indexées de courbes isoparamétriques et l'analyse des formes du nez en comparant leurs courbes correspondantes (voir figure 2.13). Les auteurs ont présenté des résultats pour les algorithmes de calcul géodésiques, le calcul des moyennes statistique et le regroupement stochastique pour effectuer la classification hiérarchique. D'autres méthodes géométriques sont basées sur la sélection des caractéristiques locales discriminantes. *Lee et al* [92] ont présenté une méthode basée sur l'extraction de huit points caractéristiques invariants du visage et leurs caractéristiques relatives. Ces caractéristiques sont directement appliquées à des algorithmes de reconnaissance du visage basés sur la programmation dynamique et le classificateur SVM. *Bronstein et al* [93, 94] ont traité les surfaces faciales comme des objets déformables dans la géométrie de Riemann et supposent que les expressions faciales peuvent être modélisées comme une surface du visage isométrique. La tâche de sélection d'une expression du visage invariant est donc transformée à une tâche de trouver une représentation d'isométrie invariante du visage. Sur cette base, ils ont proposé une représentation d'expression invariante du visage en utilisant la flexion de formes canoniques invariantes.

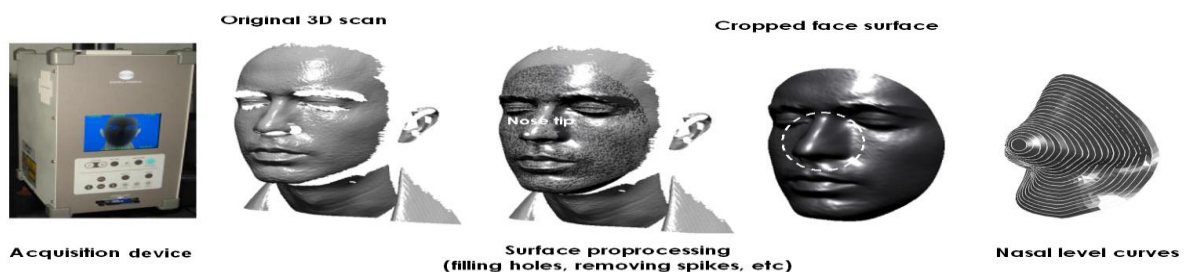


Figure 2.13 Extraction des courbes nasales [91]

Di Huang et al [95] ont proposé une nouvelle représentation géométrique pour les visages 3D afin de renforcer la distinction des images de profondeur basée sur la multi-échelle étendue LBP (ELBP) associée à la correspondance des caractéristiques locales en utilisant la méthode

proposée SIFT (Scale-Invariant Feature Transform). Cette méthode est aussi robuste aux variations d'expressions faciales.

2.6.5 Approches 2D+3D

Ces approches bimodales utilisent la combinaison des données faciale 2D et 3D pour améliorer la précision et la robustesse des systèmes de reconnaissance. Le principe de base de ces méthodes est la fusion entre l'information visuelle (généralement la couleur) et l'information 3D (modèle 3D ou image de profondeur) durant le processus de reconnaissance. On peut considérer quatre stratégies de fusion selon le niveau auquel elles interviennent, la fusion des données brutes, la fusion des descripteurs, la fusion des scores et la fusion de décision. Pour plus de détails concernant les niveaux de la fusion au sein du processus de la reconnaissance de visage, on peut consulter les deux références [96, 97]. Récemment, plusieurs approches basées sur ce principe ont été développées. *Chang et al* [73] ont étudié les améliorations possibles de la combinaison de deux systèmes de reconnaissance de visage 2D et 3D. La méthode Eigenfaces est utilisée séparément sur les images d'intensité et les images profondeur 3D. Ensuite, la fusion des données 2D et 3D est considérée au niveau du score. Les expérimentations ont été effectuées sur un ensemble de données de 275 sujets en utilisant une seule sonde et un ensemble de multi probe. Les résultats montrent que la méthode basée sur la donnée 3D surpasse celle basée sur les images 2D. Cependant, la stratégie de la fusion donne les meilleurs résultats.

Stan Z. Li et al [98] ont proposé une méthode systématique pour la fusion de données 2D et 3D pour la reconnaissance faciale à deux niveaux : au niveau de caractéristiques (données brutes) et au niveau de décisions. Les auteurs ont utilisé les caractéristiques de l'histogramme LBP pour représenter les visages 2D et 3D. Ensuite, l'algorithme Adaboost est utilisé pour la sélection des caractéristiques et pour l'apprentissage du classificateur.

Xu et al [14] ont exploité à la fois les images de profondeur 3D et les images l'intensité 2D avec les ondelette de Gabor afin de construire un classificateur robuste pour la reconnaissance de visage sous différentes variations d'expression et des poses. Un nouveau schéma de sélection hiérarchique intégré dans l'algorithme de l'analyse discriminante linéaire (LDA) et l'apprentissage AdaBoost est proposé pour sélectionner les caractéristiques les plus efficaces et les plus robustes pour construire un bon classificateur. Le système proposé par *Xu et al* est présenté dans la figure 2.14.

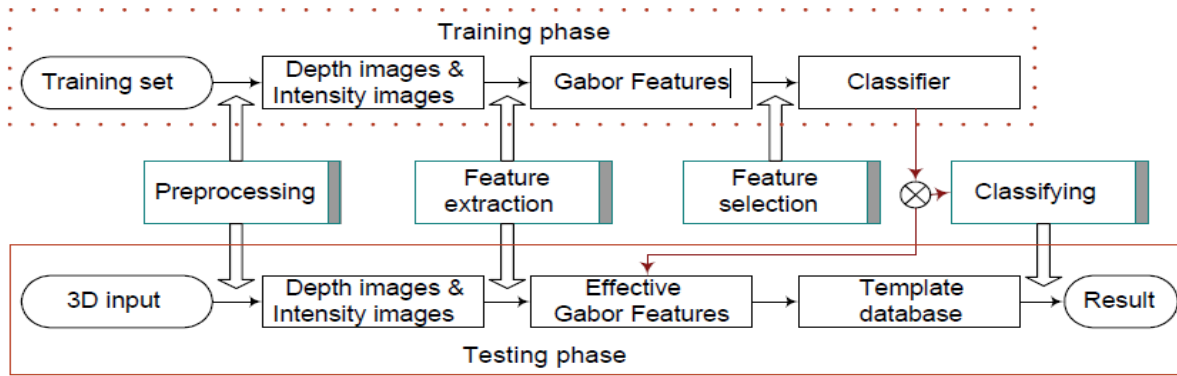


Figure 2.14 Vue d'ensemble de la méthode proposée par Xu et al [14]

Mian et al [38] ont développé une approche multimodale 2D+3D pour la reconnaissance automatique du visage, efficace aux variations d'expressions faciales. Les auteurs utilisent une nouvelle méthode de représentation sphérique de visage (3D Spherical Face Representation, 3D SFR) en combinaison avec le descripteur SIFT pour créer un classificateur de rejet, ce qui élimine rapidement un grand nombre de candidats visages au début pour une reconnaissance efficace dans le cas d'une grande galerie. Ils ont segmenté les images 3D en deux régions : le nez et le front-yeux qui sont relativement moins sensibles aux expressions. Ensuite, la correspondance de façon indépendante par la méthode ICP modifiée est effectuée. La faiblesse de cette méthode est que les auteurs considéraient la zone des yeux qui n'est pas sensible aux variations d'expressions faciales, mais en fait, le contraire est vrai.

Notre classification des approches de la reconnaissance faciale 3D est illustrée et résumée dans la figure 2.15.

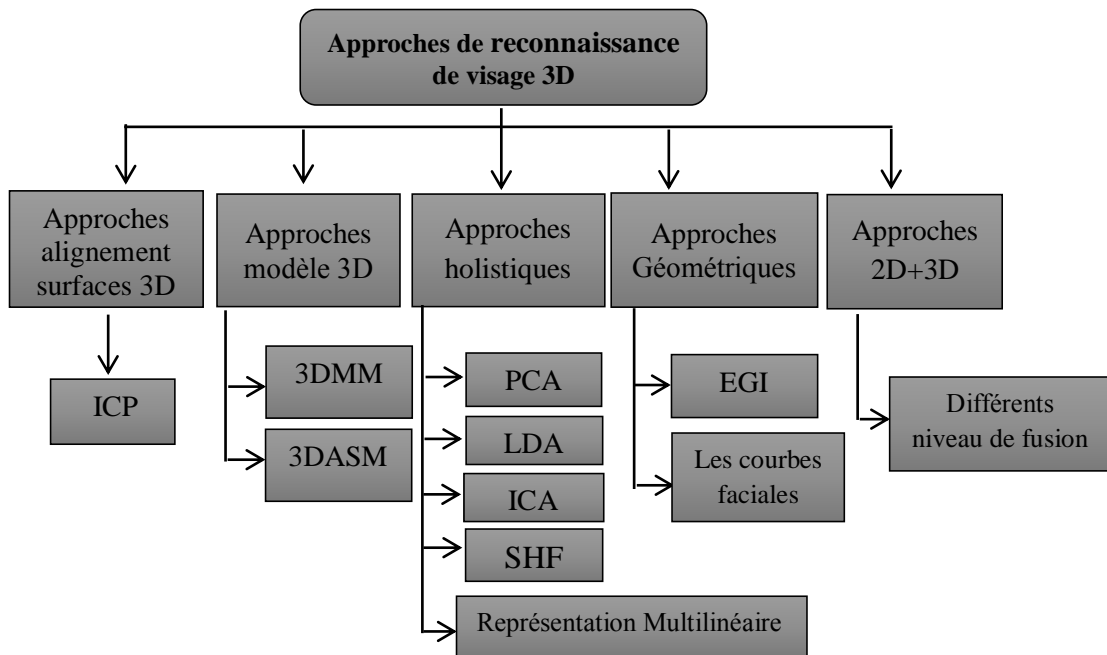


Figure 2.15 Classification des approches de la reconnaissance de visage 3D

2.7 Conclusion

Dans ce chapitre, un aperçu des techniques et des méthodes de la détection et la reconnaissance de visage 3D ont été présenté. Nous avons commencé par le passage de l'utilisation des données 2D vers le 3D et les avantages des systèmes de reconnaissance faciale 3D. Dans un premier temps, nous avons présenté en revue des méthodes de détection de visage et nous avons focalisé sur les approches 3D qui sont généralement basées sur le bout du nez comme un point de référence. Ensuite, nous avons présenté les techniques d'acquisition de données faciale 3D (les techniques actives et les techniques passives). Par la suite, nous avons mis en évidence les trois types de données les plus utilisées dans la littérature pour la représentation faciale sous forme tridimensionnelle. La bibliographie montre que les modèles basés sur les informations de profondeur de visage doivent être moins sensibles aux variations d'illuminations, d'expressions faciales et les variations de poses. Notre analyse nous a conduits à subdiviser les approches de la reconnaissance de visage 3D en cinq classes : les approches d'alignement des surfaces 3D, les approches basées sur le modèle 3D, les approches holistiques, les approches géométriques et les approches 2D+3D. Les approches holistiques prouvent leur efficacité dans le cas de la présence d'expressions faciales et variations d'illuminations.

Dans le cadre du principal objectif de notre travail, nous envisageons d'examiner le rôle de l'information de profondeur dans le système de reconnaissance de visage. Sans négliger l'information intensité dérivée de l'image couleur. Ce qui justifie notre choix du modèle dans lequel nous utilisons les informations 2D et 3D basées sur les techniques holistiques pour la reconnaissance. Nous consacrons le chapitre suivant pour l'étude de ce modèle.

Chapitre 3

Reconnaissance 2D et 3D de Visage à base d'IPC

3.1 Introduction	42
3.2 Description de la méthode proposée	42
3.3 L'alignement du visage et l'enregistrement	43
3.4 La segmentation de la zone du visage et le prétraitement	44
3.4.1 Les images de profondeur et d'intensité	44
3.4.2 La segmentation de la zone du visage basée sur IPC	45
3.4.3 Prétraitements	46
3.5 La méthode combinée de réduction de dimension PCA+EFM	48
3.6 Classification	50
3.6.1 Classification par mesure de distance	50
3.6.2 Classification par SVM	50
3.7 Résultats expérimentaux	51
3.7.1 Résultats de l'étape de détection	52
3.7.2 Résultats du système de reconnaissance	55
3.8 Conclusion	60

3.1 Introduction

Après avoir présenté les différentes méthodes de la reconnaissance faciale dans les chapitres précédents et en se basant sur les avantages et les limites de chacune d'elles. Nous nous orientons vers les méthodes globales qui préservent implicitement les informations locales sur le visage. Nous choisissons d'extraire les informations pertinentes par : la modalité profondeur et intensité qui pourrait nous être très utile pour surmonter le défi poses, la détection par IPC basée sur le bout du nez comme point de repère, les descripteurs qui sont des approches d'extraction des plus performantes dans le contexte de reconnaissance faciale, la méthode d'analyse en composantes principales (PCA) pour garder les informations les plus importantes, la méthode d'analyse discriminante linéaire améliorée pour discriminer les informations extraites et accélérer le processus d'apprentissage qui crée un modèle de classification (classificateur).

Dans ce chapitre, nous présentons donc un système de reconnaissance automatique de visage en présence d'illuminations, d'expressions et de variation de poses, basé sur l'information profondeur et intensité en utilisant la segmentation de la zone du visage avec les courbes de projection intégrale (IPC-based Facial Area Segmentation) horizontale et verticale. Nous terminerons ce chapitre par la présentation et la discussion des résultats obtenus en deux parties : les résultats de l'étape de la détection et les résultats du système de reconnaissance de visage 3D.

3.2 Description de la méthode proposée

La méthode de l'intensité maximale est utilisée pour définir le point de bout du nez. Généralement, ce point correspond à la plus grande valeur de profondeur. Cependant, il y a un problème avec certaines données inutiles tels que : les épaules, les cheveux sur le front, le cou et certaines parties de vêtements. Pour faire face à ce problème, nous proposons l'utilisation des courbes de projection intégrale (IPC) pour segmenter et extraire la zone du visage. Après cela, la méthode combinée PCA avec le modèle Fisher améliorée (EFM) sont utilisées pour obtenir les vecteurs de la matrice de caractéristiques. Ensuite, la classification est effectuée en utilisant deux méthodes : la mesure de distance et la machine à vecteurs de support (SVM). Enfin, la décision est prise de rejeter ou d'accepter la personne selon un protocole spécifique. Les expériences sont implémentées sur deux bases de données de visage 3D, CASIA3DV4 et GavabDB. Nos résultats montrent que la méthode proposée permet d'obtenir une haute performance.

Notre système de reconnaissance automatique de visage contient deux phases, la phase d'apprentissage et la phase de test. Pendant la phase d'apprentissage (off-line), l'enrôlement des images faciales est réalisé et les données d'apprentissage sont préparées pour le classificateur. La phase off-line est effectuée une seule fois par le SRV. Dans la phase de test (on-line) les nouvelles données de test sont comparées aux données d'apprentissage. Les étapes de traitement dans les deux phases sont les mêmes. La figure 3.1 illustre les différentes étapes de notre système.

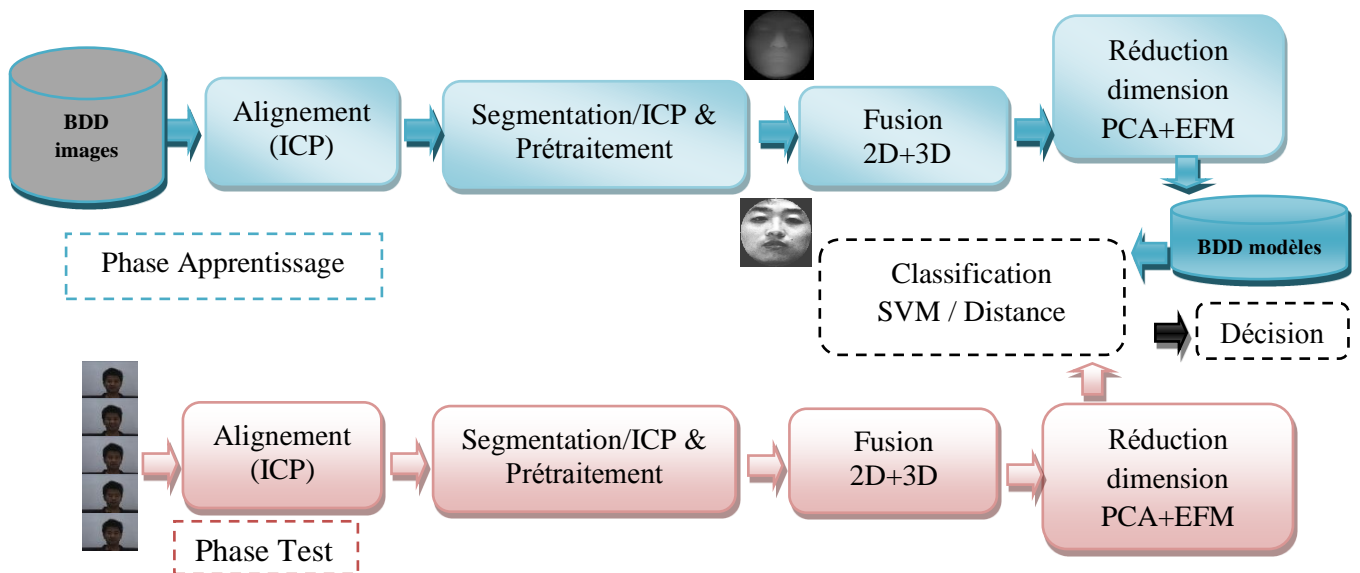


Figure 3.1 Organigramme de notre système de reconnaissance

3.3 L'alignement du visage et l'enregistrement

Dans notre travail, l'algorithme ICP est utilisé pour l'enregistrement des visages 3D. L'ICP permet de recalculer deux surfaces par itérations. Nous utilisons un modèle de visage de référence avec une expression neutre et pose frontale (0°). Tous les visages sont alignés sur ce modèle pour affiner l'effet de la variation de poses et d'expressions. L'ICP est un algorithme itératif qui minimise l'erreur quadratique moyenne entre deux ensembles de nuages de points 3D (visage de référence et visage d'entrée) en utilisant une transformation rigide. À partir d'une transformation initiale, les deux étapes suivantes sont répétées :

- *Mise en correspondance* : l'appariement de chaque primitive du modèle d'entrée avec la primitive la plus proche dans le modèle de référence.
- *Transformation rigide* : Rotation + Translation qui est généralement calculée au sens des moindres carrés.

3.4 La segmentation de la zone du visage et le prétraitement

Notre processus de détection comprend trois étapes principales : la génération des images de profondeur et d'intensité, la localisation de la zone du visage avec IPC et le prétraitement ainsi que la détection de bout du nez. Le diagramme de processus de détection proposé est présenté dans la figure 3.2.

3.4.1 Les images de profondeur et d'intensité

Avec le développement des équipements de capture 3D, il est devenu plus rapide et plus facile d'obtenir les informations de la forme 3D et de la texture 2D. Le nuage de points 3D est l'entrée de notre système. Chaque point P_i dans le nuage de points porte les coordonnées (x_i, y_i, z_i) . L'image de profondeur et l'image d'intensité sont obtenues à partir des données 3D enregistrées avec l'algorithme ICP. Comme mentionné précédemment, l'image du visage en 3D est représentée par l'image de profondeur qui est aussi appelée l'image 2.5D. Cette dernière n'est autre que la projection de l'image 3D sur le plan (X, Y) qui contient au maximum une valeur de profondeur (direction z) pour tous les points. L'image d'intensité contient l'information de la couleur sous format RGB, cette image est transformée à une image de niveaux de gris. La figure 3.3 illustre un exemple de nuage de points 3D alignés avec ICP, l'image de la profondeur 3D et l'image intensité 2D dans la base de données CASIA3DV4.

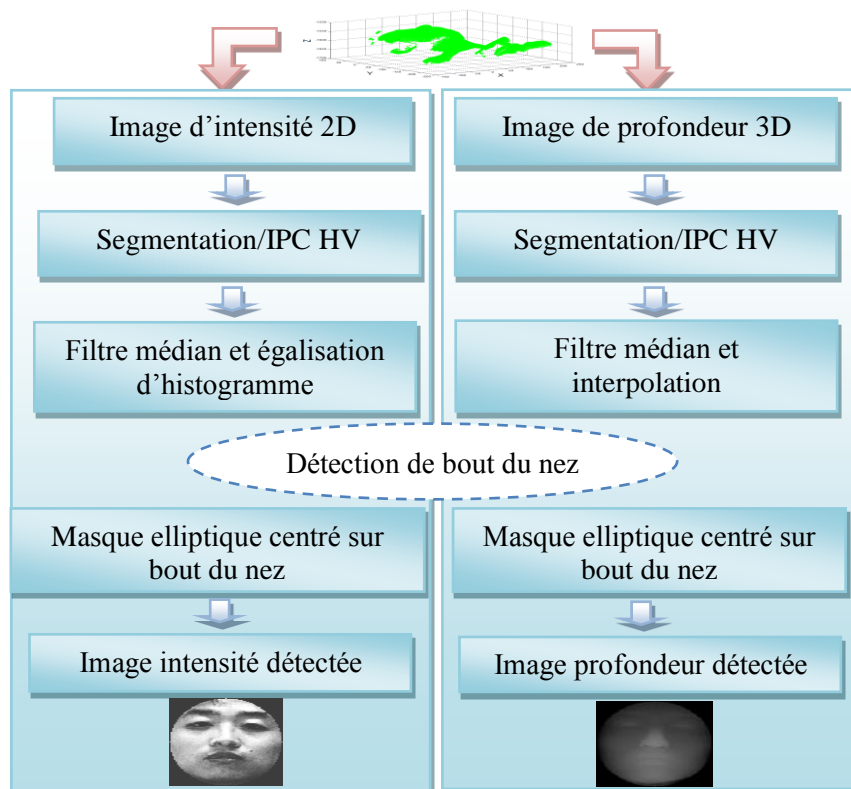


Figure 3.2 Détection de visage basée sur IPC

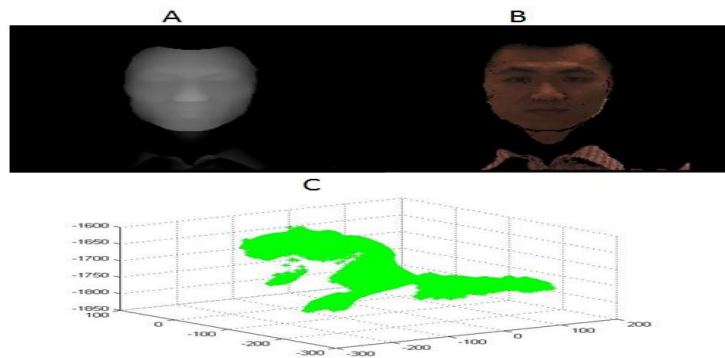


Figure 3.3 L'information de visage dans CASIA3DV4 ;
 A : image de profondeur 3D, B : image d'intensité 2D (couleur),
 C : nuage de points 3D

3.4.2 La segmentation de la zone du visage basée sur IPC

La première étape dans le système de reconnaissance de visage est de détecter la région faciale depuis l'arrière-plan. Quelques informations supplémentaires comme les épaules, les cheveux, le cou et une certaine partie de vêtements existent sur l'image. Ces données inutilisables doivent être enlevées pour obtenir une performance de détection élevée. La projection intégrale est un modèle unidimensionnel, obtenue par la somme d'un ensemble donné de pixels le long d'une direction donnée. Notre méthode est basée sur l'utilisation de l'IPC verticale et horizontale afin de segmenter la région de visage avec précision. La localisation de certains traits du visage comme les sourcils, les yeux, le nez et la bouche est réalisée grâce à l'analyse des courbes de projection d'intégrale [99]. Tout d'abord, l'image d'entrée doit être transformée afin d'être une image binaire à l'aide de la méthode de binarisation par seuillage, *Otsu* [100]. La méthode d'*Otsu* est une méthode de calcul de seuil automatique pour la binarisation des images en deux classes, le premier-plan et le fond (l'arrière-plan). L'algorithme *Otsu* suppose que l'image d'entrée contient deux classes de pixels, et leur variance intra-classe soit minimale puis on calcule le seuil optimal qui sépare ces deux classes afin que leur variance intra-classe soit minimale. Lorsque la valeur du seuil est calculée, elle est utilisée pour la binarisation de l'image. Supposant que l'image binarisée est notée par $I(x,y)$, les IPCs horizontaux et verticaux sont définis comme suit :

$$HP(y) = \sum I(\bullet, y) \quad (3.1)$$

$$VP(x) = \sum I(x, \bullet) \quad (3.2)$$

Où $HP(y)$ est la courbe de projection horizontale et $VP(x)$ est la courbe de projection verticale. Les IPCs verticales et horizontales de l'image de profondeur sont présentées dans la figure 3.4. Pour la figure 3.4 (b), l'axe X représente les coordonnées y de l'image et l'axe Y représente l'intégrale de la valeur de profondeur. Pour la figure 2.4 (c), l'axe X représente les coordonnées

x et l'axe Y représente l'intégrale de la valeur de profondeur. L'IPC horizontale donne exactement l'emplacement (x_{min}, x_{max}) de la zone du visage le long de l'axe X de l'image d'entrée. L'IPC verticale donne exactement l'emplacement (y_{min}, y_{max}) de la zone du visage le long de l'axe-Y. En se basant sur le rectangle M défini par $[x_{min} \dots x_{max}, y_{min} \dots y_{max}]$, la zone faciale est segmentée comme le montre la figure 3.4 (d). Nous pouvons voir que notre méthode de segmentation de la zone de visage basé sur IPCs est une méthode efficace. Nous l'utilisons pour supprimer les informations supplémentaires ainsi que pour segmenter la zone du visage de son arrière-plan.

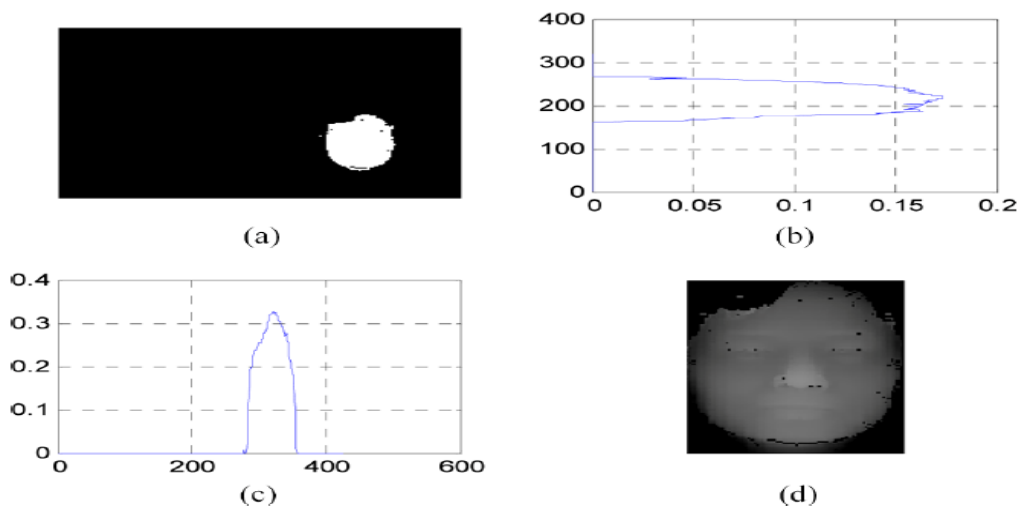


Figure 3.4 Processus de détection faciale basée sur IPC ;

(a) binarisation de l'image de profondeur, (b) projection intégrante horizontale, (c) projection intégrante verticale, (d) zone du visage segmentée

La plupart des systèmes de reconnaissance de visage proposés dans la littérature sont basés sur une intervention manuelle pour l'étape de détection. *Notre méthode basée sur les IPCs est purement automatique.*

3.4.3 Prétraitements

Après la localisation de la zone de visage comme nous l'avons mentionné précédemment. Les coordonnées du point de bout du nez sont estimées en utilisant la méthode d'intensité maximale (Maximum Intensity Method, MIM). Le nez est la partie la plus distincte dans le visage humain car c'est la plus proche au scanner 3D. Par conséquent, la zone nasale porte la plus grande valeur de profondeur. Pour chaque pixel, la moyenne des valeurs dans une fenêtre de taille 3 x 3 est calculée autour de lui. La plus grande valeur calculée dans cette fenêtre est considérée comme la région nasale. Le pixel central de cette fenêtre représente le point de bout du nez. Le prétraitement est utilisé pour améliorer l'image afin d'être vraiment expressive à la forme du visage. D'une part, un filtre médian est utilisé c'est un outil efficace pour la

suppression de bruit [101]. D'autre part, l'interpolation linéaire des pixels voisins est utilisée pour remplir les trous qui sont généralement présents dans les images de profondeur. Les images d'intensité sont traitées en utilisant un procédé similaire à celui sur les images de profondeur. L'égalisation d'histogramme est ensuite appliquée pour réduire l'effet de la variation d'illuminations. Le processus d'égalisation d'histogramme redistribue les occurrences des valeurs intensités dans l'image d'entrée pour les faire apparaître de manière plus équitable dans l'image de sortie.

La plupart des travaux sur la reconnaissance faciale utilisent un modèle elliptique pour recadrer la zone faciale et ignorent les régions en dehors de ce masque elliptique [14]. Nous l'utilisons donc et le centre de ce masque n'est autre que le bout du nez qui a été localisé précédemment. La détection et le prétraitement des images de profondeur et d'intensité sont illustrés dans les figures 3.5 et 3.6 respectivement.

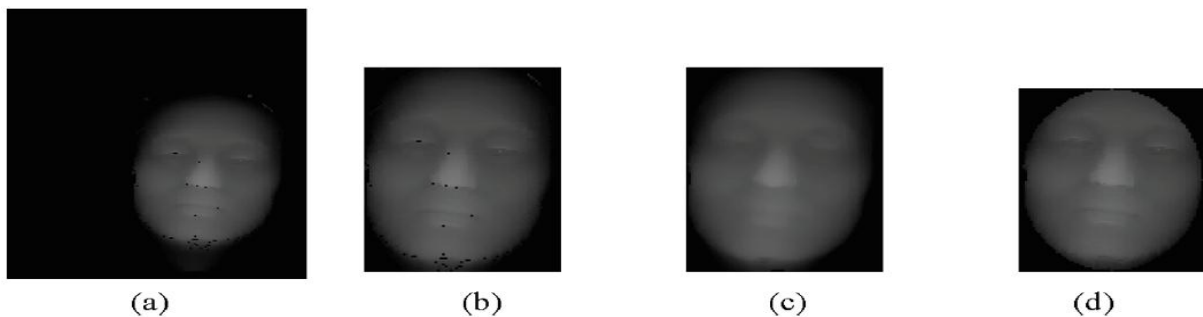


Figure 3.5 Détection et prétraitement de l'image de profondeur ;
 (a) image de profondeur d'entrée, (b) segmentation de la zone du visage par IPCs,
 (c) élimination du bruit et remplissage des trous, (d) image de profondeur après masque elliptique

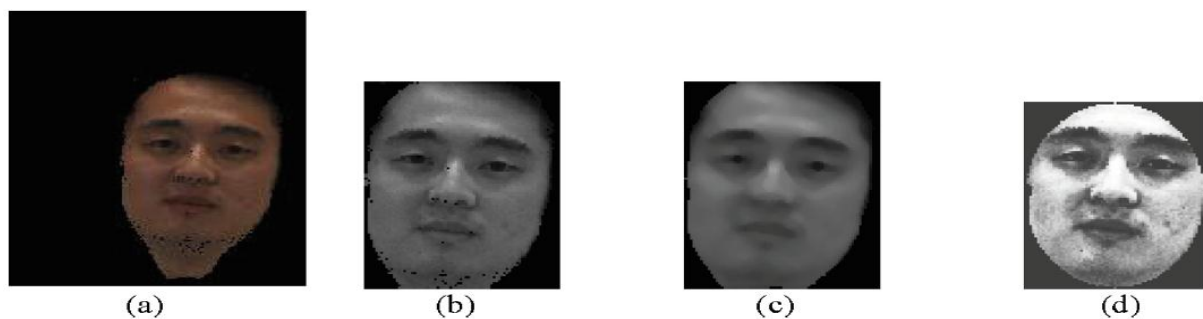


Figure 3.6 Détection et prétraitement de l'image intensité ;
 (a) image intensité d'entrée, (b) segmentation de la zone du visage par l'IPC,
 (c) élimination de bruit et remplissage des trous, (d) image d'intensité après l'égalisation d'histogramme et masque elliptique

La fusion de l'information ou des différents ensembles de caractéristiques augmente la précision des systèmes biométriques [102]. Dans ce travail, l'information d'intensité (2D) et l'information de profondeur (3D) sont fusionnées pour obtenir une meilleure représentation de

l'image du visage. Nous utilisons la fusion par combinaison basée sur les opérations logiques simples (somme produit, min, max et moyenne).

3.5 La méthode combinée de réduction de dimension PCA+EFM

La PCA est l'un des algorithmes les plus efficaces qui a été largement utilisé pour l'extraction des caractéristiques et la réduction de dimensionnalité. Cet algorithme a été proposé par *Turk et Pentland* [103] en 1991. L'analyse en composantes principales est utilisée pour projeter les images de visage de l'espace de l'image originale sur un sous-espace de visage, dans lequel la dimensionnalité est réduite tout en gardant l'information la plus importante. Tout d'abord, nous utilisons l'algorithme PCA pour réduire la dimension des vecteurs caractéristiques. Ensuite, la PCA est suivi par un algorithme étendu de LDA qui est nommé EFM, afin d'obtenir une caractéristique vectorielle de petite dimension et d'augmenter la puissance de la discrimination dans l'espace des caractéristiques.

Un algorithme qui résume la méthode combinée PCA + EFM est présenté ci-dessous. L'image d'entrée (une classe) est représentée par un seul vecteur, on considère une matrice $A = [A_1 A_2 \dots A_M]$, avec N lignes et M colonnes, M est le nombre des images d'apprentissage et N est la taille du vecteur caractéristique. Chaque classe ou chaque personne est représentée par un nombre de vecteurs colonnes avec différentes variations d'illuminations, d'expressions et d'expressions sous une variation d'illuminations selon le protocole décrit dans le tableau 3.1. La matrice A est définie comme suit :

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1M} \\ A_{21} & A_{22} & \dots & A_{2M} \\ \dots & \dots & \dots & \dots \\ A_{N1} & A_{N2} & \dots & A_{NM} \end{bmatrix}$$

Tout d'abord, l'algorithme PCA est utilisé pour trouver une matrice de transformation linéaire U_{PCA} de chaque vecteur de caractéristique dans le sous-espace des vecteurs propres. Où W_i (équation 3.6) est le vecteur caractéristique d'apprentissage projeté dans le sous-espace de vecteurs propres. Les étapes pour calculer la matrice U_{PCA} peuvent être résumées comme suit :

- *Étape 1* : Trouver la moyenne de chaque visage \bar{A} , où $A_i (i = 1, 2, \dots, M)$ représente le $i^{\text{ème}}$ vecteur colonne de A ;

$$\bar{A} = \frac{1}{M} \sum_1^M A_i \quad (3.3)$$

- *Étape 2* : Soustraire le visage moyenne \bar{A} de chaque visage d'apprentissage

$$Q_i = A_i - \bar{A} \quad (3.4)$$

- *Étape 3* : Calculer la matrice de covariance C à partir de la nouvelle matrice X , où $X = [Q_1, Q_2, \dots, Q_M]$

$$C = X X^T \quad (3.5)$$

- *Étape 4* : Calculer les valeurs propres V et les vecteurs propres U de la matrice C et trier les vecteurs propres dans l'ordre décroissant. La matrice U_{PCA} contient les k premiers vecteurs propres correspondant aux k plus grandes valeurs propres. Nous utilisons la matrice de transformation linéaire globale U_{PCA} pour la projection. Nous trouvons la projection de chaque vecteur caractéristique dans le sous-espace des vecteurs propres, sous la forme de :

$$W = U_{PCA} * A \quad (3.6)$$

Deuxièmement, afin d'améliorer la puissance de la discrimination entre les classes (vecteur colonne), l'EFM est utilisée après la PCA. Les vecteurs d'apprentissage projetés dans le sous-espace des vecteurs propres de l'algorithme PCA (W) sont des entrées pour l'EFM, grâce à ce processus, les caractéristiques les plus importantes sont extraites et préparées pour la classification. L'algorithme EFM est présenté ci-dessous :

- *Étape 5* : Trouver les matrices de dispersion : intra-classes (S_w) et inter-classes (S_b)

$$S_w = \sum_{i=1}^M \sum_{j=1}^{n_i} (W_{ij} - \bar{m}_i)(W_{ij} - \bar{m}_i)^T \quad (3.7)$$

$$S_b = \sum_{i=1}^M n_i (\bar{m}_i - \bar{m})(\bar{m}_i - \bar{m})^T \quad (3.8)$$

Où : W_{ij} est le $j^{\text{ième}}$ des échantillons de la classe, \bar{m}_i est la moyenne des échantillons dans la classe i , \bar{m} est la moyenne de tous les échantillons, n_i est le nombre des échantillons dans la classe i .

- *Étape 6* : Calculer les valeurs propres (Y) et les vecteurs propres (E) de la matrice S_w
- *Étape 7* : Calculer la nouvelle matrice inter-classes K_b :

$$K_b = Y^{-1/2} E^T S_b E Y^{-1/2} \quad (3.9)$$

- *Étape 8* : Calculer les valeurs propres (O) et les vecteurs propres (H) de la matrice K_b .
- *Étape 9* : Calculer la matrice de transformation globale U_{EFM}

$$U_{EFM} = E Y^{-1/2} H \quad (3.10)$$

Enfin, la matrice d'apprentissage final (W_{final}) projeté dans le sous-espace est calculée en utilisant la matrice de transformation linéaire globale U_{EFM} , sous la forme de :

$$W_{final} = U_{EFM} * W \quad (3.11)$$

La projection des images de tests est assurée par le même processus et comparée avec chaque projection des images d'apprentissage au niveau du processus de la classification selon le protocole utilisé.

3.6 Classification

Après la représentation de chaque image de visage avec un grand pouvoir discriminant, la classification est la dernière étape dans notre système où la décision est réalisée entre le visage de test et les visages d'apprentissages stockés dans la base de données. Dans cette sous-section, se basant sur les travaux [96] et [97] nous nous intéressons à deux types de classifications: la première est basée sur la mesure de similarité en utilisant la distance corrélation normalisée et la seconde est basée sur le classificateur SVM.

3.6.1 Classification par mesure de distance

Comme la distance euclidienne et la distance de Mahalanobis, la corrélation (aussi appelée : corrélation normalisée, coefficient de corrélation ou la similarité de cosinus) est une mesure largement utilisée dans les systèmes de reconnaissance de formes [104, 105] pour décrire la similarité S entre deux vecteurs A et B . La distance de corrélation s'est avérée être la meilleure distance performante par rapport aux autres distances [96, 105]. La corrélation S entre deux vecteurs, A et B est définie par :

$$S(A, B) = \frac{A^T \cdot B}{\|A\| \cdot \|B\|} \quad (3.12)$$

Si on trouve une valeur élevée de corrélation c.à.d. on a une bonne similarité entre les deux vecteurs A et B et vice-versa.

3.6.2 Classification par SVM

Le SVM est une nouvelle technique d'apprentissage statistique utilisée pour l'analyse des données et la reconnaissance des formes, proposé par *Cortes et Vapnik* [106]. L'algorithme SVM a été développé au cours des années 1990 à des fins industrielles [105, 107]. Il peut traiter nombreux problèmes tels que la classification, la régression et la fusion. Au début et dans sa forme de base, le SVM est utilisé comme une méthode de classification binaire basé sur un problème à deux classes. Le SVM binaire cherche à trouver l'hyperplan de séparation optimale entre les deux classes en maximisant la marge entre l'hyperplan et les deux classes qui sont étiquetées avec -1 et 1. Supposons que A est un ensemble de données, x_i ($i = 1, 2, \dots, K$) sont les vecteurs caractéristiques d'apprentissage en k -dimension et y_i sont les étiquettes (labels) :

$$A = \{(x_i, y_i) | x_i \in R^k, y_i \in \{-1, +1\}\} \quad (3.13)$$

Pour le SVM linéaire, l'hyperplan de séparation optimale peut être exprimé par la fonction suivante :

$$f(x) = (w \cdot x) + b \quad (3.14)$$

Pour le SVM non linéaire, la fonction de décision est une fonction non linéaire. Les données d'entrée sont reconstruites dans un espace de dimension élevée basé sur une fonction de noyau (kernel) afin d'augmenter la précision de la classification. La précision du classificateur SVM dépend de la fonction kernel utilisée par celui-ci [108]. La fonction kernel la plus populaire et est la fonction Gaussienne ou la fonction à base radiale (Radial Basis Function, RBF). Cette dernière surpasse les noyaux linéaires ainsi que les noyaux polynomiaux dans les applications des systèmes de reconnaissance de visage [61, 109, 110]. Dans la plupart des problèmes réels il n'y a pas de séparation linéaire possible entre les données. Le classificateur de marge maximale ne peut pas être utilisé car il fonctionne seulement si les classes de données d'apprentissage sont linéairement séparables [36] (voir la figure 3.7).

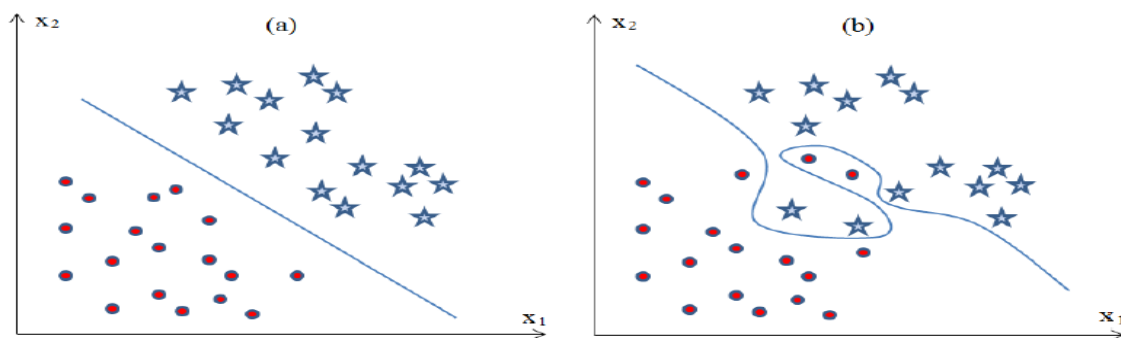


Figure 3.7 Classification SVM ;
 (a) SVM linéaire séparation par une ligne droite,
 (b) SVM non linéaire séparation par une courbe [36]

Le SVM est généralisé pour résoudre le problème multi-classes. Les algorithmes SVM multi-classes peuvent être divisés en deux catégories : One-Versus-All et One-Versus-One [110, 111]. Lorsque le nombre de classes les personnes dans notre système de reconnaissance est assez grand, nous utilisons la stratégie One-Versus-All basé sur le noyau RBF pour effectuer la vérification des images faciales entre les imposteurs et les clients. One-Versus-All est une méthode simple dans laquelle nous utilisons M classificateurs, un pour chaque classe. Les M classificateurs sont combinés pour prendre la décision finale.

3.7 Résultats expérimentaux

Dans cette section, nous présentons les résultats expérimentaux du système proposé en deux parties : les résultats de l'étape de détection et les résultats du système de reconnaissance. Notre système de reconnaissance est appliqué sur deux bases de données de visage CASIA3DV4 et GavabDB.

Entre août et septembre 2004, la base de données de visage CASIA3DV4 est collectée. Elle est constituée de 4624 scans de 123 personnes utilisant le scanner 3D Minolta Vivid 910. La BDD CASIA3DV4 est caractérisée par de nombreuses variations complexes qui sont difficiles à valider pour beaucoup d'algorithmes. Nous avons utilisé 20 images pour chaque individu (personne) dans deux protocoles différents avec un nombre total de 2168 images. Comme point de départ, nous allons étudier l'influence d'illuminations et les variations d'expressions puis les variations de poses dans la deuxième expérience. Le nombre total des images de profondeur utilisée est de 2168 images où :

- les images 1, 2, 3, 4 et 5 représentent des variations d'illuminations : la lumière de bureau, éclairage de haut, éclairage de bas, éclairage à gauche et à droite ;
- les images 5, 6, 7, 8 et 9 représentent des variations d'expressions : rire, sourire, colère, surprise et œil fermé ;
- les images 10, 11, 12, 13, 14 et 15 contiennent une combinaison de variation d'expressions sous différentes illuminations ;
- les images 16, 17, 18, 20, 21 représentent des variations de poses : pose frontale (0°), rotation à droite ($20-30^\circ$), rotation à gauche ($20,30^\circ$), rotation à droite ($50-60^\circ$) et rotation à gauche ($50-60^\circ$) respectivement ;
- l'image 19 représente le profil de chaque personne, l'angle de rotation est de 90° . Dans notre travail, cet angle n'est pas pris en compte.

Pour la base de données GavabDB, le scanner utilisé pour l'acquisition est le Minolta VI-700. Il y a 61 personnes différentes ayant 9 images pour chaque personne. Nous prenons en compte 7 modèles pour chaque personne, le nombre total des images de profondeur utilisée est de 427 images ou :

- les images 1 et 2 sont des vues frontales avec une expression neutre ;
- les images 3 et 4 sont des vues frontales avec des variations d'expression, rire et sourire ;
- les images 5 et 6 sont des vues avec une variation de pose de $\pm 35^\circ$, regardant vers le haut et vers le bas ;
- l'image 7 est un geste aléatoire choisi par l'utilisateur.

3.7.1 Résultats de l'étape de détection

La base de données CASIA3DV4 (format WRL) contient 123 personnes différentes ayant 37 ou 38 individus (modèles). Pour chaque scan, une image d'intensité 2D et une image de profondeur 3D sont générées. Dans CASIA3DV4, nous considérons plusieurs variations des

poses, d'expressions, d'illuminations, de variations combinées d'expressions sous illuminations et des poses sous expressions (voir figure 3.8, 3.9 et 3.10). Pour démontrer l'efficacité de notre méthode de détection, nous utilisons 15 modèles pour chaque personne : 5 images avec des variations d'illumination, 5 images avec des variations d'expressions (rire, sourire, colère, surprise, œil fermé) et 5 images avec des variations d'expressions sous différentes variations d'illuminations. La seconde base de données utilisée est GavabDB, elle contient 549 images tridimensionnelles de surfaces faciales. Dans cette base de données, il y a 61 individus différents avec 9 images pour chaque personne. Chaque image est donnée par un maillage de nuage de points 3D sans texture, contrairement à la base de données CASIA3DV4 qui contient les images d'intensité. La GavabDB comporte plusieurs variantes concernant les poses et les expressions faciales. Les 9 images correspondantes à chaque individu (voir figure 3.11) sont : deux vues frontales avec une expression neutre, deux vues avec une rotation de la tête 35° (regardant vers le haut en regardant vers le bas respectivement) sous une expression neutre, deux images avec une rotation de la tête 90° (les profils gauche et droit) sous une expression neutre, 2 images frontales avec des variations d'expression (rire et sourire) et une seule image avec une geste aléatoire choisi par l'utilisateur.



Figure 3.8 Les variations d'expression dans la base de données CASIA3DV4



Figure 3.9 Les variations d'illuminations dans la base de données CASIA3DV4



Figure 3.10 Les variations de poses dans la base de données CASIA3DV4



Figure 3.11 Exemples des 9 vues pour la même personne dans GavabDB

Dans l'étape de détection, cinq modèles pour chaque individu sont utilisés : deux images frontales avec une expression neutre, deux images frontales avec expressions (rire et sourire) et une seule image avec le geste aléatoire choisi par l'utilisateur. Les résultats de l'étape de détection dans les deux bases de données CASIA3DV4 et GavabDB sont représentés dans le tableau 3.1 et 3.2 respectivement.

Tableau 3.1 La précision de la détection dans CASIA3DV4

Dataset	Nombre de personnes	Nbr d'images par personne	Correctement détectées		Incorrectement détectées		Précision de détection (%)	
			D	I	D	I	D	I
IV	123	5	605	609	10	6	98,37	99,02
EV	123	5	607	608	8	7	98,69	98,86
EVI	123	5	592	597	23	18	96,26	97,07

Tableau 3.2 La précision de la détection dans GavabDB

Dataset	Nbr de personnes	Nbr d'images par personne	Correctement détectées	Incorrectement détectées	Précision de détection (%)
A	61	2	116	6	95,08
B	61	2	114	8	93,44
C	61	2	55	6	90,16

Pour le **tableau 3.1**, on a :

- *IV (615 images) : Variations d'illumination ; (y compris l'éclairage haut, bas, gauche et droite)*
- *EV (615 images) : Variations d'expression ; (y compris sourire, rire, colère, surprise et yeux fermés)*
- *EVI (615 images) : Variation d'expressions sous des variations d'illuminations.*
- *D (Depth) : Image de profondeur 3D et I (Intensity) : Image d'intensité 2D.*

Pour le **tableau 3.2**, on a :

- *A : Deux vues frontales avec une expression neutre.*

- *B* : Deux vues frontales avec des variations d'expressions (rire, sourire).
- *C* : Un geste aléatoire choisi par l'utilisateur.

En se basant sur les résultats présentés dans les tableaux 3.1 et 3.2, nous pouvons formuler les conclusions suivantes :

- La précision de détection élevée de 99,02% est obtenue avec une variation d'illuminations, comme indiqué dans le tableau 3.1. nous pouvons affirmer que la méthode proposée est très efficace en présence de variation d'illuminations.
- Le taux le plus bas obtenu est de 90,16% dans le tableau 3.2 avec la base de données GavabDB. Dans ce cas, il y a un geste aléatoire choisi par l'utilisateur, il peut être une orientation de la tête ou un obstacle devant la face ce qui a conduit à une diminution de précision.
- Il convient de souligner que les images d'intensité sont plus robustes que les images de profondeur en présence d'illumination dans l'étape de détection, cela est dû à l'effet d'égalisation d'histogramme qui est appliquée à ces images lors de l'étape de prétraitement.
- Le point de référence dans notre méthode de détection est le bout du nez qui a la plus grande valeur de profondeur. Une limitation de l'approche proposée se produit lorsque certaines personnes ont la valeur la profondeur du menton ou des cheveux sur le front plus grande que la valeur de la profondeur du nez. Ce problème diminue la précision de la détection.

3.7.2 Résultats du système de reconnaissance

Cette sous-section décrit les résultats expérimentaux du système de reconnaissance. Il y a trois parties dans cette étude pour examiner la performance des informations 2D et 3D dans le système de reconnaissance de visage. Les bases de données du visage doivent comporter plusieurs défis et variations. Ce qui nous permet de tester plusieurs techniques et algorithmes pour résoudre ces nombreuses difficultés et problèmes. Comme on le verra les expressions, l'illumination et les variations de poses dans les images d'apprentissage et les images de tests sont très différentes pour chaque individu.

Dans la première expérience, nous étudions l'influence d'illuminations et d'expressions avec CASIA3DV4 selon le protocole 1 présenté dans le tableau 3.3 (phase 1). La seconde expérience est réalisée en utilisant la base de données GavabDB afin d'évaluer la stabilité et la performance de notre système en présence de faibles variations de pose (regarder vers le haut et le bas avec $\pm 35^\circ$). Enfin, notre système est testé en présence d'une grande variation de poses (regarder gauche et droite ($\pm 20-30^\circ$, $\pm 50-60^\circ$)) avec la base de données CASIA3DV4 selon le protocole 2 présenté dans le tableau 3.3 (phase 2).

Tableau 3.3 Les protocoles utilisés pour la base de données CASIA3DV4

Dataset	Protocole 1 (phase 1)		Protocole 2 (phase 2)	
	Clients	Imposteurs	Clients	Imposteurs
Apprentissage	500 images (1, 4, 8, 9, 10)	0 images	600 images (1, 4, 8, 9, 10,16)	0 images
Evaluation	500 images (2, 6, 7, 14, 15)	195 images (1:15)	600 images (2, 6, 7, 14, 15,17)	208 images (1:15+18)
Test	500 images (3, 5, 11, 12, 13)	150 images (1:15)	600 images (3, 5, 11, 12, 13,20)	160 images (1:15+21)

Dans la première expérience, les 123 personnes du CASIA3DV4 sont séparées en deux classes, les clients et les imposteurs. La classe client contient 100 personnes et la classe imposteur contient 23 personnes subdivisées en 13 imposteurs pour l'évaluation et 10 imposteurs pour le test. Dans l'ensemble d'apprentissage, nous avons 100 personnes avec 5 images (1, 4, 8, 9, 10) comme des clients, cet ensemble ne contient pas d'imposteurs (0 image).

Dans l'ensemble d'évaluation, la classe client contient les mêmes 100 personnes de l'ensemble d'apprentissage, mais avec 5 autres conditions (2, 6, 7, 14, 15). La classe imposteur dans l'ensemble d'évaluation contient 13 personnes avec toutes les 15 images (1:15).

Dans l'ensemble de test, la classe client contient les mêmes 100 personnes de l'ensemble de l'apprentissage, mais avec 5 autres conditions (3, 5, 11, 12, 13). La classe imposteur dans l'ensemble de test contient 10 personnes avec toutes les 15 images (1:15). Suivant le même protocole pour la deuxième expérience sur CASIA3DV4, mais dans le protocole 2 nous ajoutons les différentes variantes de poses (16, 17, 18, 20, 21) pour l'ensemble des données.

La deuxième expérience est réalisée en utilisant la base de données GavabDB. Les 61 personnes de la base de données GavabDB sont subdivisées en deux classes, client et imposteur. La classe client contient 55 personnes et la classe imposteur contient 6 personnes. Dans l'ensemble d'apprentissage, nous prenons 4 images pour chaque personne (les images 1, 3, 5,7) comme clients. Dans l'ensemble de test, la classe client contient les mêmes 55 personnes d'apprentissage avec 3 autres conditions (les images : 2, 4, 6). La classe imposteur dans l'ensemble de test contient 6 personnes avec toutes les 7 images (les images 1: 7). Le protocole utilisé pour la base de données GavabDB est présenté dans le tableau 3.4.

Tableau 3.4 Le protocole utilisé pour la base de données GavabDB

Dataset	Client	Imposteur
Apprentissage	220 images (1, 3, 5,7)	0 image
Test	165 images (2, 4, 6)	42 images (1:7)

A travers nos résultats nous pouvons dire que :

- 1) la taille de la base de données joue un rôle important pour la stabilité du système de reconnaissance. Les figures 3.12 et 3.13 illustrent la variation de TEE et TV respectivement, en fonction du nombre des caractéristiques (EFM) avec le classificateur SVM et la mesure de la distance (DIS). Les courbes de TV et TEE sont plus stables dans la base CASIA3DV4 (2168 images) par rapport à la base GavabDB (427 images). Par exemple, nous avons obtenu pour un nombre de caractéristiques égale à 10, un TV = 81,67% sur GavabDB et un TV = 93.13% sur CASIA3DV4 lorsque nous utilisons le SVM. Lorsque nous utilisons la classification avec la mesure de distance, nous obtenons pour le même nombre de caractéristiques (10) un TV = 26,81% sur GavabDB et TV = 89,60% sur CASIA3DV4. La différence est grande (54,86 %) pour la base GavabDB et TV (3,5%) pour la base CASIA3DV4.
- 2) Pour l'orientation de la tête avec un angle inférieur ou égal à 35°, de bons résultats ont été obtenus, TV = 95,85% et TEE = 2,07%. Ces résultats sont proches de celles obtenus dans le cas de la pose frontale (angle = 0 °) avec un TV = 96,75% et TEE = 1,6%. Par conséquent, la variation de poses jusqu'à 35° n'affecte pas notre système. Nous constatons que pour les angles supérieur ou égal à 35 ° (jusqu'à 60 °), les performances diminuent un TV = 89,63% et TEE = 3,16%. Dans ce travail, l'ICP est utilisé pour l'alignement de l'angle de pose ; cet algorithme est limité à la correction des angles inférieurs à 30-35°.
- 3) Nous pouvons noter que le SVM est toujours le meilleur classificateur comme le montrent les figures 3.12 et 3.13. Il donne le meilleur TV et TEE dans toutes nos expériences avec les deux bases de données. Les résultats prometteurs obtenus dans la présence de plusieurs défis (illuminations, expressions et poses) sont dus : d'une part au bon détecteur avec la segmentation de la zone du visage basée sur l'IPC et d'autre part, grâce à l'utilisation de la PCA suivie de l'EFM dans l'étape d'extraction des caractéristiques. La PCA réduit la grande dimension de l'espace de données vers une dimension intrinsèque plus petite et l'EFM préserve un équilibre adéquat entre la sélection des valeurs propres et l'exigence que les valeurs propres de la matrice de

dispersion intra-classe ne sont pas trop petites. Le nombre de caractéristiques qui représente le nombre de lignes de la matrice de transformation globale U_{EFM} varie afin de déterminer les caractéristiques optimales. Comme le montrent les figures 3.12 et 3.13, nous observons qu'avec un petit nombre de caractéristiques (30-35) nous obtenons la meilleure performance de reconnaissance (TV et TEE) dans toutes les expériences.

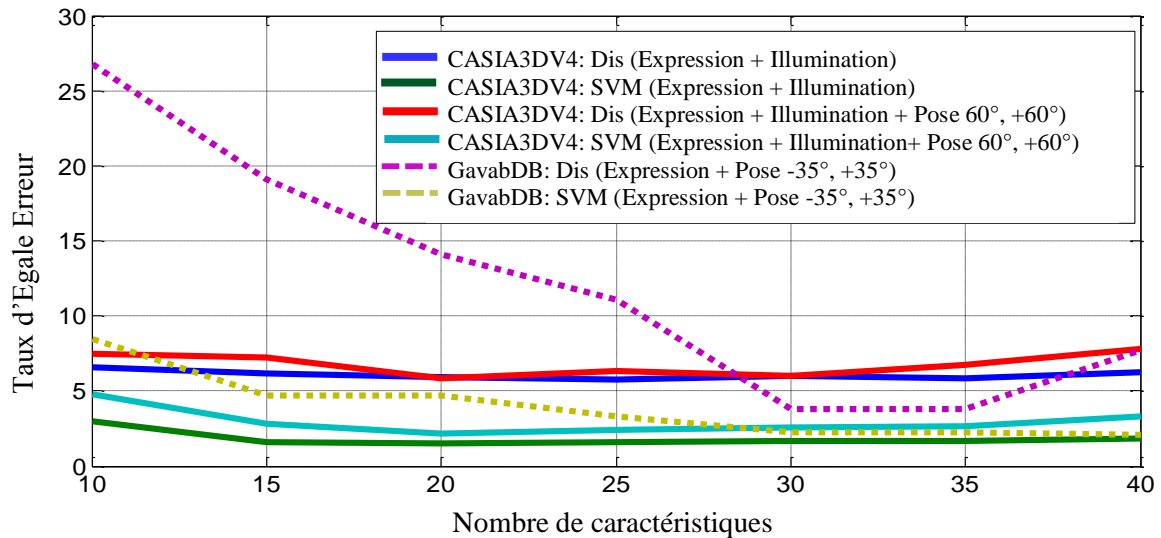


Figure 3.12 Variations de TEE en fonction du nombre de caractéristiques

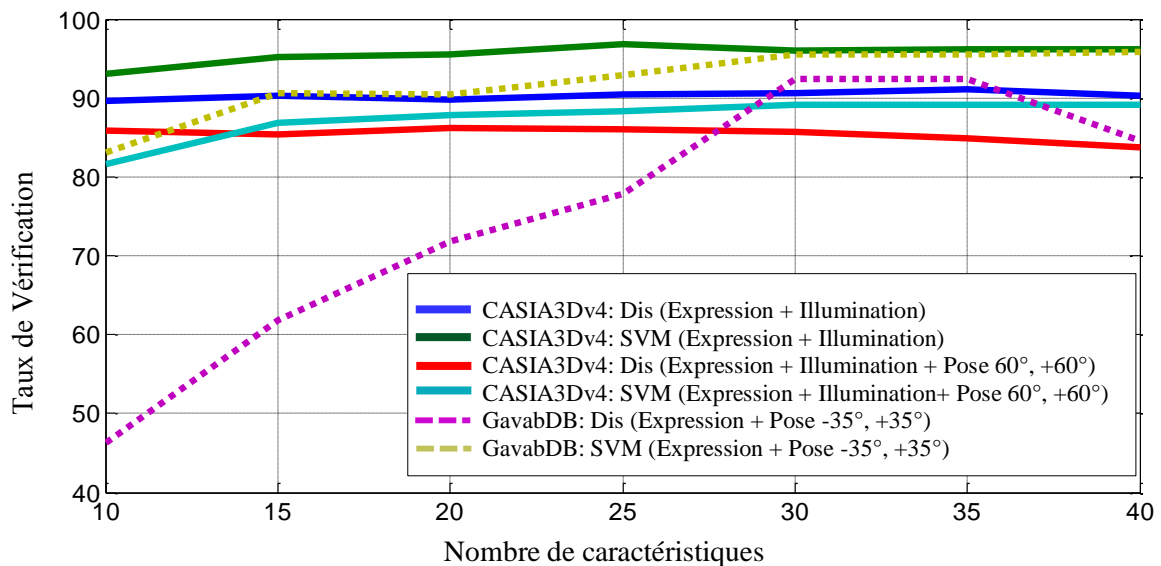


Figure 3.13 Variations de TV en fonction du nombre de caractéristiques

Nos meilleurs résultats sur les deux bases de données avec les différents défis sont résumés et présentés dans le tableau 3.5.

La comparaison des TV avec l'état de l'art est illustrée dans le tableau 3.6 pour la base GavabDB et dans le tableau 3.7 pour la base CASIA3DV4. D'après nos expériences sur les deux bases de données GavabDB et CASIA3DV4, des résultats encourageants sont atteints proches des meilleurs résultats obtenus dans l'état de l'art.

Tableau 3.5 Meilleurs résultats de reconnaissance pour différentes expériences

Défis et variations de données.	GavabDB	CASIA3DV4	
	Expressions, poses (haut/bas (35°)) et geste aléatoire.	Phase 1	Phase 2
		Illuminations, expressions et expressions sous illuminations	Illuminations, expressions, expressions sous illuminations et poses (droite/gauche (20-30°, 50-60°)).
TEE (%)	2.07	1.60	3.16
TFA (%)	0.50	1.45	3.20
TFR (%)	3.63	1.80	7.16
HTER (%)	2.06	1.65	5.18
TV (%)	95.85	96.75	89.63

Tableau 3.6 Comparaison de TV sur GavabDB

Référence	Méthodes	TV et défis
<i>Moreno et al</i> [112]	Segmentation HK, PCA et SVM	90.16%: expression neutre 77.9% : geste aléatoire et poses
<i>Xiaoxing et al</i> [113]	Représentations parcimonieuses, Caractéristiques de bas niveau, Distance géodésique	93.33% : variation d'expressions 94.68%: neutre + expressions
<i>Drira et al</i> [114]	Courbes radiaux élastiques, PCA	94.54% : variation d'expressions 94.67%: neutre + expressions
<i>Mahoor et al</i> [115]	Distance d'Hausdorff, ICP	90.16%: expression neutre 78%: neutre + expressions
<i>Mousavi et al</i> [116]	2DPCA, SVM	91%: neutre + expressions
<i>Notre méthode</i> [9]	Segmentation IPC, PCA+ EFM, SVM.	95.85%: neutre + expressions+ poses (haut/bas) et geste aléatoire.

Tableau 3.7 Comparaison de TV sur CASIA3DV4

Référence	Méthodes	TV et défis
<i>Yue et al</i> [117]	3DLBP, Régression Spectrale; NN	94.17 %: variation d'expressions
<i>Xu et al</i> [14]	Gabor, LDA, Adaboost	93.3%: expressions + illuminations 91.0%: poses 50-60°.
<i>Yong et al</i> [117]	Caractéristiques Géométriques, ICP, LDA, D Géodésique	91.10%: Galerie : 90 expressions neutres. Sonde : 2610 autres modèles
<i>Zhou et al</i> [118]	Fusion évolutif, 2D PCA, LBP, mesure de similarité cosinus	94.68% : expressions + illuminations
<i>Notre méthode</i> [9]	Segmentation du visage à base d'IPC, PCA avec EFM, SVM.	96.75% : illuminations + expressions 89.63% : illuminations + expressions + poses (20-30°, 50-60°).

3.8 Conclusion

Dans ce chapitre, nous avons proposé un système automatique de reconnaissance de visage 3D utilisant les images de profondeur et d'intensité. La région faciale est détectée par la segmentation avec IPC dans laquelle certaines informations supplémentaires et inutilisables sont éliminées. La localisation de bout du nez est assurée par la méthode d'intensité maximale pour recadrer la zone faciale à l'aide d'un masque elliptique centré sur ce point. Les images sont prétraitées. Les vecteurs caractéristiques obtenus sont extraits et réduits en utilisant une méthode efficace la PCA suivie par l'EFM. Ensuite, la mesure de distance et le SVM sont utilisés pour la classification. Les expériences sont effectuées sur les bases de données de visage CASIA3DV4 et GavabDB. La comparaison avec les méthodes existantes montrent que nos résultats sont prometteurs.

Bien que notre méthode offre une bonne robustesse pour les variations d'illuminations et d'expressions sauf qu'elle reste limitée quant aux poses dépassant 35°.

Chapitre 4

Descripteurs Locaux et Réduction de Dimension pour Vérification V3D Expressions

4.1 Introduction	62
4.2 Travaux antérieurs.....	62
4.3 Aperçu de l'approche proposée	65
4.4 Les descripteurs locaux	66
4.4.1 Local Phase Quantization (LPQ)	67
4.4.2 Local Binary Pattern (LBP)	69
4.4.3 LBP basé sur les patchs (Patch based LBPs).....	70
4.4.3.1 Three-Patch Local Binary Patterns (TPLBP).....	70
4.4.3.2 Four-Patch Local Binary Patterns (FPLBP)	72
4.4.5 Binarized Statistical Image Features (BSIF)	73
4.5 Extraction des caractéristiques d'histogrammes	75
4.6 Réduction de dimension basée sur OLPP	76
4.7 Résultats expérimentaux et discussion	78
4.7.1 Base de données et protocole de vérification	78
4.7.2 Analyse de résultats et discussion	79
4.8 Conclusion.....	84

4.1 Introduction

Dans ce chapitre, nous proposons un nouveau cadre pour la vérification du visage 3D en utilisant la réduction de dimension basée sur les descripteurs locaux dans un environnement non contrôlé. Les histogrammes des descripteurs locaux sont utilisés pour représenter distinctement les images faciales. À cet effet, différents descripteurs locaux sont évalués : LBP, TPLBP, FPLBP, BSIF et LPQ. Des expériences sur les combinaisons de ces descripteurs sont effectuées au niveau des caractéristiques à l'aide d'une concaténation simple des histogrammes. Tout système de reconnaissance biométrique comporte une phase très importante basée sur la réduction d'espace. Pour cela, nous consacrons un chapitre pour présenter les différentes techniques de projection d'espace. La classification n'est pas à négliger. Nous essayons de présenter l'essentiel des approches utilisées. La réduction de la dimension se pose comme une étape primordiale dans le processus de prétraitement des données (filtrage, nettoyage, élimination des points aberrants, etc.). En effet, pour des données appartenant à un espace de grande dimension, certains attributs n'apportent aucune information voire expriment du bruit, d'autres sont redondants ou corrélés. Ceci rend les algorithmes de décision complexes, inefficaces, moins généralisables et d'interprétation délicate. Les méthodes de réduction de la dimension de l'espace de représentation peuvent être divisées en méthodes d'extraction d'attributs et méthodes de sélection d'attributs. L'extraction d'attributs transforme l'espace d'attributs de départ en un nouvel espace formé de la combinaison linéaire ou non linéaire des attributs initiaux.

L'approche proposée est évaluée avec différents algorithmes de réduction de la dimensionnalité : PCA, OLPP et la méthode combinée PCA+EFM. Enfin, le SVM multi-classes est utilisé comme classificateur pour effectuer la vérification entre les imposteurs et les clients. La méthode proposée est testée sur la base de données CASIA3DV4 et les résultats montrent que notre méthode permet d'obtenir de hautes performances de vérification.

4.2 Travaux antérieurs

Cette section met en évidence les travaux connexes concernant la représentation d'image faciale basée sur les descripteurs locaux, ainsi que les méthodes populaires de la réduction de dimension et leurs utilisations dans les systèmes de reconnaissance faciale.

Xuebing et al [119] ont utilisé les informations de forme 3D pour développer un système de reconnaissance automatique de visage basé sur les caractéristiques d'histogramme. Le bout du nez est détecté et l'axe de symétrie du visage est déterminé en utilisant l'algorithme PCA ainsi

que des calculs de la courbure faciale. L'image faciale est subdivisée en plusieurs bandes horizontales afin d'exploiter plus d'informations sur la géométrie locale. Ensuite, les caractéristiques de l'image faciale sont extraites de la valeur de profondeur dans chaque bande, où le vecteur caractéristique correspond à l'histogramme de chaque bande. La distance euclidienne norme L2 et la corrélation sont utilisées pour comparer les vecteurs caractéristiques. La méthode proposée a été évaluée avec la base de données FRGC v2 et un TEE=5.8% est obtenu. Une autre méthode basée sur la spécification de l'histogramme local (Local Histogram Specification, LHS) est proposé par *Liu et al* [120], pour résoudre le problème de la haute fréquence et la basse fréquence des caractéristiques d'illuminations dans l'image faciale, un filtre passe-haut est appliqué pour éliminer l'illumination de basse fréquence. Les histogrammes locaux et les histogrammes locaux statistiques sont exploités à partir des images faciales avec l'illumination normale (naturel). La combinaison des descripteurs de visage LBP, l'ondelette de Gabor et le MBC (Monogenic Binary Coding) sont utilisés pour améliorer la représentation des images faciales. D'autre part, les auteurs utilisent la fusion des deux méthodes d'extraction des caractéristiques, BFLD (Block-based Fisher's Linear Discriminant) et KLDA (Kernel Linear Discriminant Analysis). Les auteurs affirment que l'amélioration significative a été atteinte grâce à la combinaison des caractéristiques hétérogènes ainsi qu'aux procédés d'extraction de caractéristiques. L'inconvénient de cette méthode est que les auteurs utilisent un découpage manuel pour la sélection de la zone faciale. Un taux de reconnaissance de 81,6% est obtenu sur la base de données de visage 2D, CAS-PEAL-R1.

Parmi les excellents descripteurs locaux utilisés pour la représentation de l'apparence des images faciales on trouve LBP, LPQ et leurs méthodes étendues. Ces descripteurs ont été largement utilisés dans les systèmes de reconnaissance de visage. Le TPLBP et le FPLBP sont de nouveaux descripteurs basés sur les patches proposés par *Wolf et al* [121], les auteurs ont comparé les performances des méthodes d'apprentissage de similitude avec les méthodes basées sur le descripteur pour l'identification de visage. Ces deux méthodes TPLBP et FPLBP sont basées sur les statistiques de patches afin d'améliorer les performances du descripteur LBP basique. Afin de renforcer la capacité de discrimination de descripteur LBP pour les images de profondeur *Amel Aissaoui et al* [122] ont proposé un nouveau descripteur appelé le DLBP (Depth LBP) qui donne une description précise et rapide de la variation de forme des images de profondeur en prenant en compte le signe et la magnitude de la différence de voisinage des pixels. Le classificateur du proche voisin est adopté pour la reconnaissance et de bons résultats sont obtenus. *Ahonen, et al* [123] ont travaillé sur la reconnaissance des visages avec des images faciales floues: l'histogramme de LPQ est calculé dans les régions locales. Les auteurs prouvent

que LPQ est très tolérant au flou. Presque avec la même idée, la Multi-échelle LPQ (MLPQ) est proposée par *Chi Ho, et al* [124], les histogrammes locaux d'un descripteur multi-échelle des images faciales floues sont projetés dans l'espace LDA afin d'augmenter les performances. La MLPQ est calculée sur de petites régions dans l'image du visage, ces caractéristiques sont combinées en utilisant la fusion par l'algorithme de l'analyse discriminante Kernel Discriminant Analysis (KDA). Par ailleurs, *Baohua, et al* [125] ont exploité le potentiel des deux descripteurs LPQ et LBP afin d'obtenir les caractéristiques les plus importantes où les vecteurs caractéristiques de descripteurs LBP et PLQ sont concaténés, c'est une fusion simple au niveau de caractéristiques. Dans ce travail, les auteurs prouvent que la méthode combinée LPQ + LBP donne des meilleurs résultats que l'utilisation de chaque méthode de façon indépendante. *Berretti et al* [126] ont proposé un système de reconnaissance faciale 3D qui subdivise le visage 3D en un ensemble des bandes isogéodésique. Puis, un descripteur nommé 3D weighted walkthroughs (3DWWs) a été utilisé pour représenter ces bandes. Ce descripteur fournit une représentation approximative de la morphologie locale de visages qui présente des variations lisses pour les changements induits par les expressions faciales. Un algorithme de correspondance à base de graphes a été utilisé pour faire l'identification.

Le BSIF est un autre descripteur excellent, récemment proposé par *Kanala et E. Rahtu* [127] pour la reconnaissance faciale et la classification de texture. Le descripteur BSIF exploite les caractéristiques statistiques de l'image basée sur l'apprentissage automatique d'un ensemble fixe de filtres à l'aide d'un petit ensemble d'images naturelles pour construire un code binaire. Ce dernier, représente efficacement les images d'entrée au niveau de pixel. Afin d'améliorer la précision, les histogrammes de codes binaires de pixels dans les régions locales sont utilisés. La recherche concernant le descripteur BSIF est encore en développement et des nouvelles idées sont constamment ajoutées.

Dans le domaine de traitement d'image et notamment pour la reconnaissance du visage, la réduction de la dimensionnalité joue un rôle très important non seulement pour les difficultés de traiter la haute dimension des caractéristiques faciales, mais aussi pour le problème de multi-classes (personnes). Pour la reconnaissance du visage, il peut y avoir plus de centaines de personnes dans la base de données, et chaque personne a des dizaines d'images, ce qui nous adonne un problème de discrimination entre les classes. On peut dire que c'est un défi de la classification multi-classes. Cela, sans oublier le problème du temps de calcul pour les données de haute dimension qui nécessite un matériel spécifique pour le traitement. Récemment, plusieurs approches de réduction de la dimensionnalité ont été proposées. L'une d'elles, est l'algorithme OLPP proposée par *C. Deng et al* [128], un nouveau cadre pour la reconnaissance

faciale est développé basé sur OLPP (un plus de détails sur cette méthode est présenté dans la section 4.6.1). Les auteurs mettent en évidence que l'algorithme OLPP présente une puissance discriminante plus que la méthode de Laplacian faces (LPP). Cependant, seules les bases de données 2D (Yale, ORL et PIE) sont utilisées pour l'évaluation expérimentale. Dans le même contexte, *L. Wang, et al* [129] ont ajouté les ondelettes de Gabor pour l'algorithme OLPP afin d'améliorer la performance de leur système de reconnaissance d'expressions faciales. Lorsque les caractéristiques de Gabor sont extraites, l'OLPP est utilisé pour réduire la dimension de ces caractéristiques qui sont classés en utilisant le SVM. Leurs expérimentations ont été implémentées sur les bases de données 2D ORL et Yale. Les auteurs de cet article ont prouvé que Gabor-OLLP dépasse les algorithmes : PCA, LDA et LPP.

4.3 Aperçu de l'approche proposée

Dans cette étude, les caractéristiques locales du visage sont utilisées pour améliorer la performance d'un système de vérification de visage 3D à partir d'images de profondeur 2.5D. Tout d'abord, la région du visage est détectée puis prétraitée. Ensuite, l'image faciale est passée à travers l'un des descripteurs locaux LBP, TPLBP, FPLBP, BSIF ensuite découpée en plusieurs blocs rectangulaires. À cet effet, l'histogramme de chaque bloc est extrait. Ensuite, les histogrammes de tous les blocs sont concaténés dans un vecteur caractéristique unique. Ces vecteurs sont caractérisés par une grande dimensionnalité. Afin de réduire la dimension de ces vecteurs, différentes méthodes, notamment PCA, OLPP et PCA+EFM sont utilisées. Enfin, le SVM multi-classes est adopté pour la classification et la vérification. L'approche proposée se divise en deux phases : la phase d'apprentissage (l'enrôlement) et la phase de test. Nous utilisons la base de données de visage CASIA3DV4, qui contient 123 personnes avec différentes variations d'illumination, d'expressions et illumination combinée avec des expressions. Les sujets (personnes) sont formés en tant que clients et imposteurs selon un protocole spécifique au cours de la phase d'apprentissage. Le même traitement est effectué dans les deux phases de manière à obtenir les vecteurs caractéristiques.

Dans la phase de test, les vecteurs caractéristiques sont appariés avec chaque échantillon défini dans la phase de l'apprentissage à l'aide de SVM multi-classes afin d'accepter la personne de test en tant que client ou la rejeter comme imposteur. L'organigramme de l'approche proposée est montré dans la figure 4.1. Généralement, le prétraitement est un processus nécessaire pour le domaine de traitement d'image. L'objectif du processus de prétraitement est l'amélioration de l'effet visuel des images faciales et leur clarté afin d'être vraiment expressive

de la forme de visage humain. Notre méthode de prétraitement et détection est décrite en détail dans le chapitre 3, section 3.4. Cependant, dans ce chapitre, nous utilisons les images de profondeur 2.5D comme entrée pour le système RV3D. La figure 4.2 montre un exemple de prétraitement de l'image de profondeur.

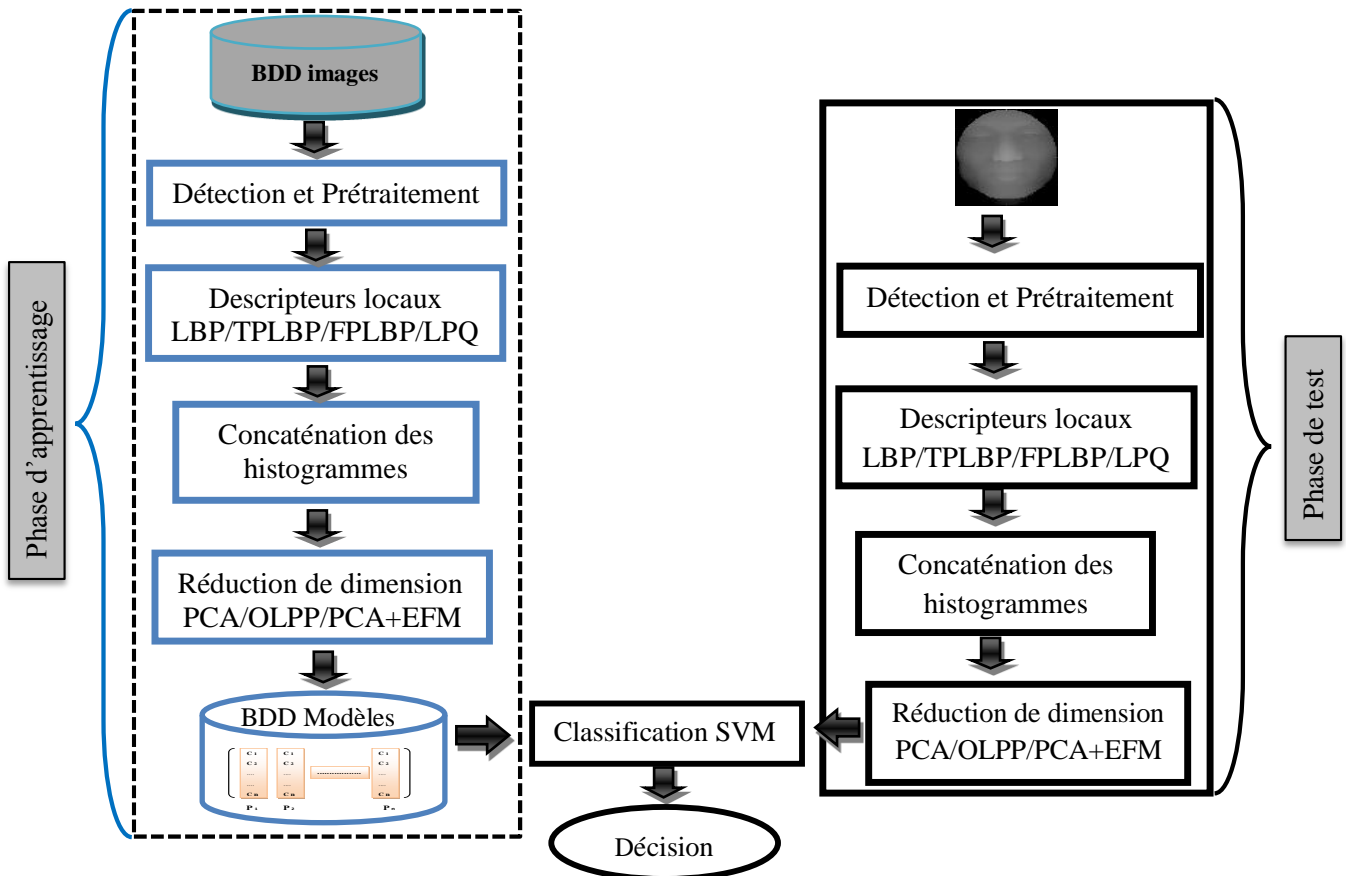


Figure 4.1 Organigramme de l'approche proposée

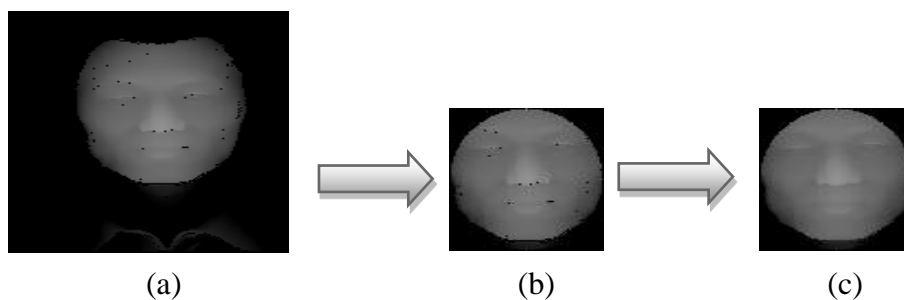


Figure 4.2 Prétraitement de l'image de profondeur 3D ;
 (a) Image d'entrée, (b) Image détectée, (c) Image prétraitée

4.4 Les descripteurs locaux

La représentation de l'image faciale est un processus important pour l'efficacité du système de vérification. Dans lequel, l'information appropriée est obtenue pour être invariante et discriminante. Les descripteurs locaux ont été considérés comme l'une des méthodes les plus efficaces adoptées à cet effet.

Les descripteurs locaux peuvent être utilisés dans la reconnaissance faciale, l'analyse de texture, la reconnaissance d'objets et la détection des régions d'intérêt [130]. Récemment, ces méthodes ont été largement utilisées pour les systèmes de reconnaissance de visage [122, 130-137]. L'idée principale des descripteurs locaux est de représenter les images faciales de manière discriminative avec leurs micros caractéristiques d'apparence locale.

4.4.1 Local Phase Quantization (LPQ)

La quantification de la phase locale ou le descripteur LPQ a été désigné pour la première fois par *Ojansivu et Heikkilä* [138] pour l'utiliser dans la classification de textures pour les images floues. Il permet d'améliorer la classification de textures pour être robuste aux artéfacts générés par le flou présent dans une image [139]. Le descripteur LPQ est construit de façon à ne retenir dans une image que l'information locale invariante à un certain type de flou. Il est insensible au flou central symétrique, tel que celui causé par le mouvement linéaire et hors du foyer du capteur [138]. Inspiré par cette idée, nous proposons le descripteur LPQ comme une méthode efficace pour résoudre le problème des variations d'expressions dans le système de vérification du visage 3D basé sur les images de profondeur. Dans notre système les images faciales comprennent des mouvements dans différentes régions causées par différentes expressions telles que le rire, le sourire, la colère, la surprise etc..., l'opérateur LPQ est basé sur la transformée de Fourier de la phase. L'extraction de l'information de la phase locale est utilisée par l'application de la transformée de Fourier à court terme (STFT) calculée sur un rectangle N_x de $M * M$ voisins pour chaque pixel x dans l'image faciale $f(x)$ définie par l'équation (4.1), où w_u correspond aux vecteurs de base de la décomposition à la fréquence u , f_x contient toutes les valeurs de l'image appartenant au voisinage N_x .

$$F(u, x) = \sum_{v \in x} f(x - y) e^{2j\pi u^T y} = w_u f_x \quad (4.1)$$

La transformée de Fourier est alors calculée pour seulement 4 fréquences u_i ($i = 1, \dots, 4$): $u_1 = [a, 0]^T$, $u_2 = [0, a]^T$, $u_3 = [a, a]^T$ et $u_4 = [a, -a]^T$, où a représente la fréquence scalaire suffisamment élevée pour $H_{u_i} > 0$, alors nous obtenons un vecteurs F_x^c où :

$$F_x^c = [F(u_1, x), F(u_2, x), F(u_3, x), F(u_4, x)] \quad (4.2)$$

Par la suite, un quantificateur scalaire simple est utilisé pour l'extraction des informations de la phase dans chaque coefficients de Fourier en observant les signes des parties réelles (*Re*) et imaginaires (*Im*). Le quantificateur scalaire est donné par l'équation suivante :

$$q_j(x) = \begin{cases} 1 & g_j(x) \geq 0 \\ 0 & \text{ailleurs} \end{cases} \quad (4.3)$$

Où $g_j(x)$ représente la $j^{\text{ième}}$ composante du vecteur $G_x = [Re\{F_x\}, Im\{F_x\}]$. Les huit coefficients binaires obtenus $q_j(x)$ sont représentés comme des valeurs entières entre 0 et 255 en utilisant un codage binaire simple pour obtenir les étiquettes de LPQ, F_{LPQ} qui est définie par :

$$F_{LPQ}(X) = \sum_{i=1}^8 q_i(x) 2^{i-1} \quad (4.4)$$

En conséquence, nous obtenons l'étiquette d'image F_{LPQ} , dont les valeurs sont invariantes pour le flou (barbouillage). La figure 4.3 présente l'organigramme de l'ensemble des étapes nécessaires à la construction du descripteur LPQ pour une image faciale. La figure 4.4 illustre une représentation de l'image de profondeur 3D par le descripteur LPQ en voisinage de pixels $Nz = 3, 5, 7$ et 9 .

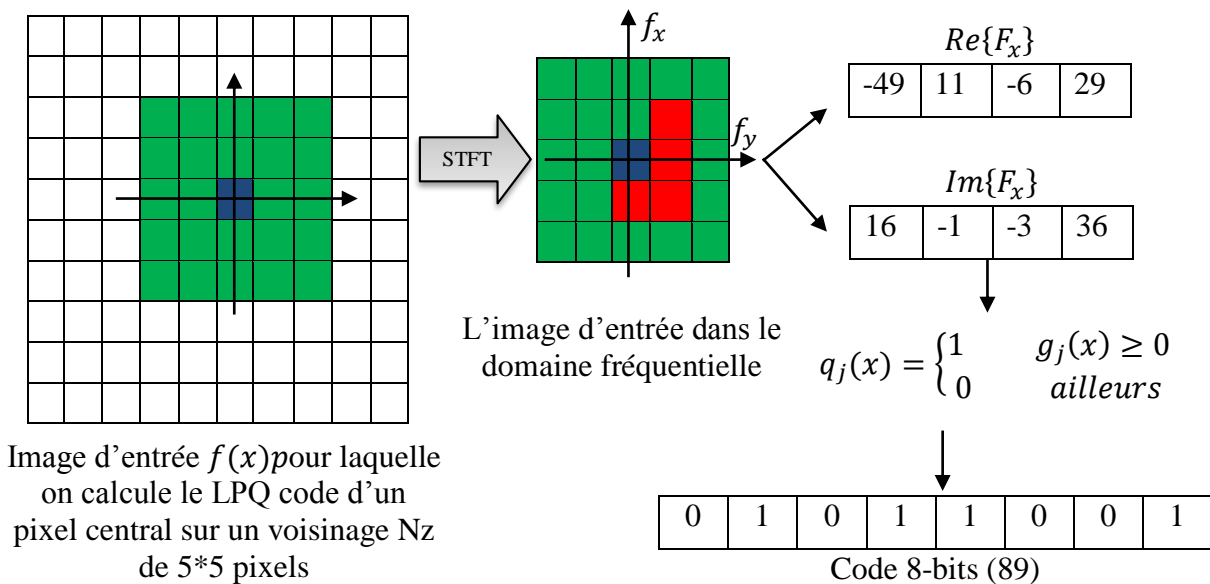


Figure 4.3 Organigramme de l'ensemble des étapes nécessaires du descripteur LPQ

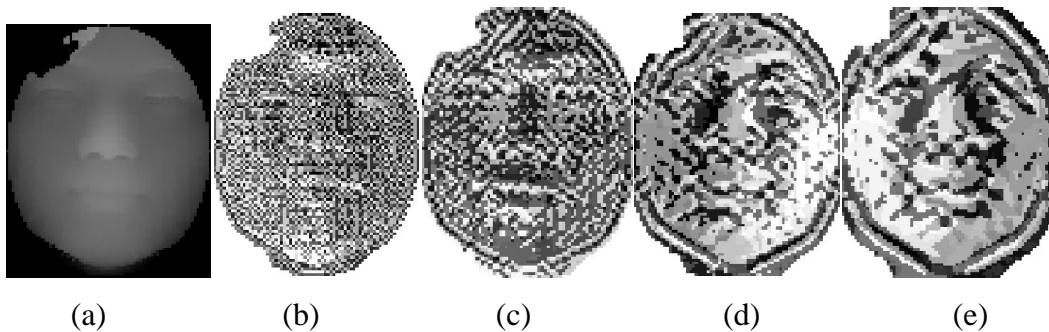


Figure 4.4 Représentation d'une image de profondeur avec le descripteur LPQ sous différents voisinages de pixel Nz ;
 (a) image d'entrée, (b) code LPQ ($Nz=3$), (c) code LPQ ($Nz=5$), (d) code LPQ ($Nz=7$),
 (e) code LPQ ($Nz=9$)

4.4.2 Local Binary Pattern (LBP)

L'opérateur LBP a été initialement proposé par *T. Ojala et al* [140] en 2002 afin d'exprimer la texture des patches de l'image. Il a été largement appliqué avec divers algorithmes de systèmes de reconnaissance de visage comme une méthode d'extraction de caractéristiques locales [125]. L'opérateur LBP de base attribue à chaque pixel un motif binaire. Le LBP d'une image de pixel est produit par le seuillage du voisinage 3 * 3 avec le pixel central (si la valeur du pixel central était supérieure à la valeur du son pixel voisin ou non) et le transfert comme un code binaire qui est converti en un nombre décimal. La figure 4.5 donne une illustration de l'opérateur LBP de base. Après cela, cet opérateur a été étendu pour utiliser de différents rayons de voisinages R et différents points d'échantillonnage P [141] ce qui permet d'extraire les caractéristiques dans différentes échelles. Soit I_c un pixel dans l'image d'entrée, ses pixels voisins sont $(I_0, I_1, \dots, I_{P-1})$. La réponse LBP du pixel I_c est calculée comme suit :

$$LBP = \sum_{P=0}^{P-1} s(I_P - I_c) 2^P \tag{4.5}$$

$$s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} \tag{4.6}$$

Le LBP étendu sélectionne les pixels voisins comme un ensemble de points d'échantillonnage réparti uniformément le long d'un cercle avec comme centre le point I_c (pixel central) et un rayon R comme représenté dans la figure 4.6.

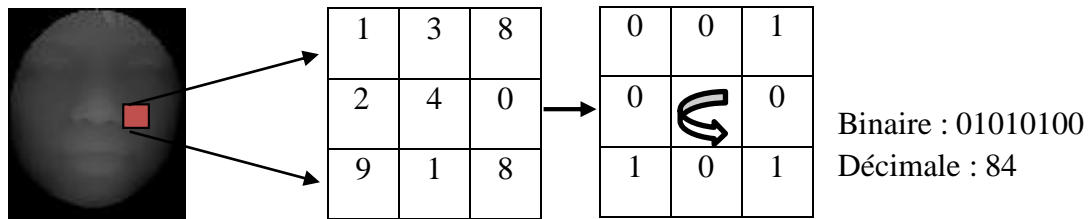


Figure 4.5 Une illustration de LBP basique

Dans la littérature la notion LBP est généralement utilisée pour désigner l'opérateur LBP basique, tandis que la notion $LBP_{P,R}$ est utilisée pour représenter l'étendue LBP où, l'indice P représente le nombre des points d'échantillonnage et l'indice R représente le rayon du cercle.

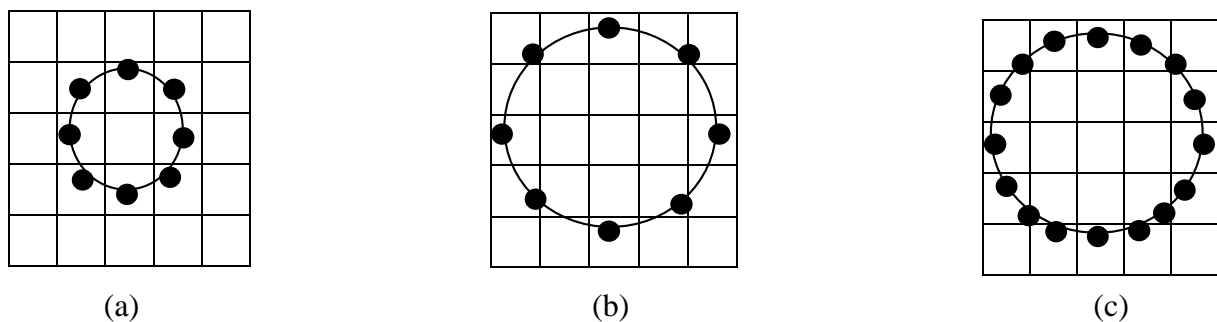


Figure 4.6 Exemples de d'opérateur $LBP_{P,R}$; (a) $LBP_{8,1}$, (b) $LBP_{8,2}$, (c) $LBP_{16,2}$

Le nombre total des valeurs des différentes sorties est 2^P qui sont générées par l'opérateur $LBP_{P,R}$ avec certaines valeurs correspondantes aux mêmes motifs suivant la rotation. Afin d'éliminer l'effet de la rotation et pour améliorer encore la puissance discriminante du LBP d'autres méthodes étendues sont proposées, telle que le Rotation-invariant LBP ($LBP_{P,R}^{ri}$) et l'Uniform-rotation-invariant LBP ($LBP_{P,R}^{uri}$) [22, 140, 142]. La figure 4.7 illustre une représentation d'image de profondeur avec le descripteur LBP et différentes valeurs de P et R .

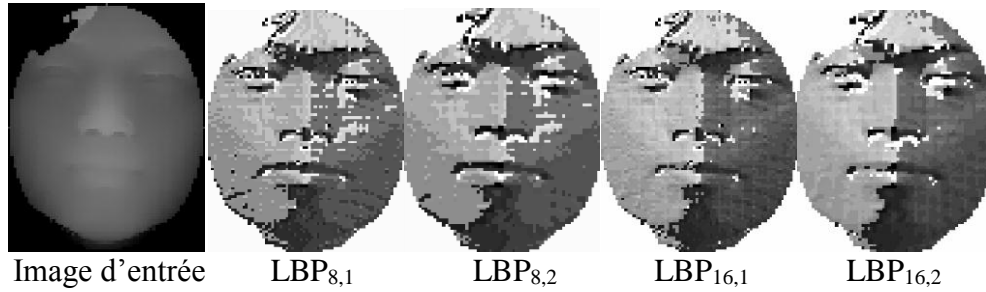


Figure 4.7 Représentation d'une image de profondeur avec le descripteur LBP

4.4.3 LBP basé sur les patches (Patch based LBPs)

Les descripteurs LBP basés sur les patches sont placés dans une famille de nouveaux descripteurs d'images introduites en 2008 par *Wolf et al* [121]. Au lieu de faire la comparaison entre les pixels voisins et un pixel central pour donner à chaque pixel son motif, la comparaison ici se fait entre les patches voisins et un patch central. Deux types de ces descripteurs ont été proposés, Three-Patch Local Binary Patterns (TPLBP) et Four-Patch Local Binary Patterns (FPLBP).

4.4.3.1 Three-Patch Local Binary Patterns (TPLBP)

Le descripteur TPLBP produit une valeur d'un seul bit dans le code attribué à chaque pixel à travers une comparaison de la similarité entre trois patches différents. La similarité entre les patches est calculée en utilisant la mesure de distance avec la norme L2. Pour chaque pixel de l'image, on considère un patch de taille $w * w$, centré dans ce pixel. Les patches voisins sont distribués uniformément dans un anneau de rayon r autour du patch central (voir figure 4.8). Ce dernier est comparé avec les deux patches qui se situent entre une valeur définie α . La valeur d'un seul bit est définie en fonction de celle des deux patches la plus similaire. Le descripteur TPLBP est donné par l'équation suivante :

$$TPLBP = \sum_{i=0}^{S-1} f(d(P_0, P_i) - d(P_0, P_{i+\alpha \bmod(S)}))2^i \quad (4.7)$$

L'exemple ci-dessous explique le codage du pixel selon le descripteur TPLBP avec $\alpha=2$, $S=8$, et $w=3$:

$$\begin{aligned}
 TPLBP = & f(d(P_0, P_1) - d(P_0, P_3))2^0 + f(d(P_0, P_2) - d(P_0, P_4))2^1 \\
 & + f(d(P_0, P_3) - d(P_0, P_5))2^2 + f(d(P_0, P_4) - d(P_0, P_6))2^3 \\
 & + f(d(P_0, P_5) - d(P_0, P_7))2^4 + f(d(P_0, P_6) - d(P_0, P_8))2^5 \\
 & + f(d(P_0, P_7) - d(P_0, P_1))2^6 + f(d(P_0, P_8) - d(P_0, P_2))2^7
 \end{aligned}$$

Où d est la mesure de distance L_2 entre les patches, P_0 patch central, P_i les patches voisins du patch central ($i = 1, \dots, 8$) et f est la fonction de seuillage qui est calculée par :

$$f(x) = \begin{cases} 1 & \text{si } x \geq \tau \\ 0 & \text{si } x < \tau \end{cases} \quad (4.8)$$

τ : Le seuil de comparaison est choisi légèrement plus grand que 0 (dans notre cas=0.01) afin d'assurer une certaine stabilité dans les régions uniformes [121, 143]. La figure 4.9 illustre une représentation de l'image de profondeur avec le descripteur TPLBP. La rangée du haut de gauche à droite : l'image de profondeur d'entrée, le code image TPLBP (S=8 et r=3) avec $w=3, 5, 7$ et 9. La rangée du bas le code image TPLBP (S=8 et r=5) avec $w=3, 5, 7$ et 9.

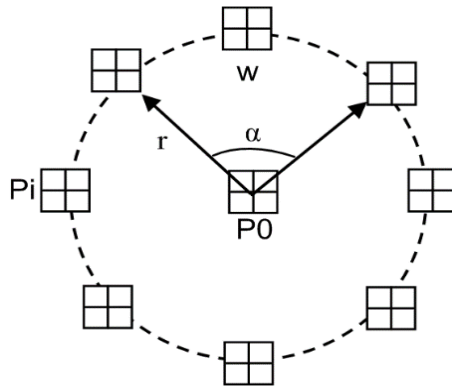


Figure 4.8 La procédure de descripteur TPLBP

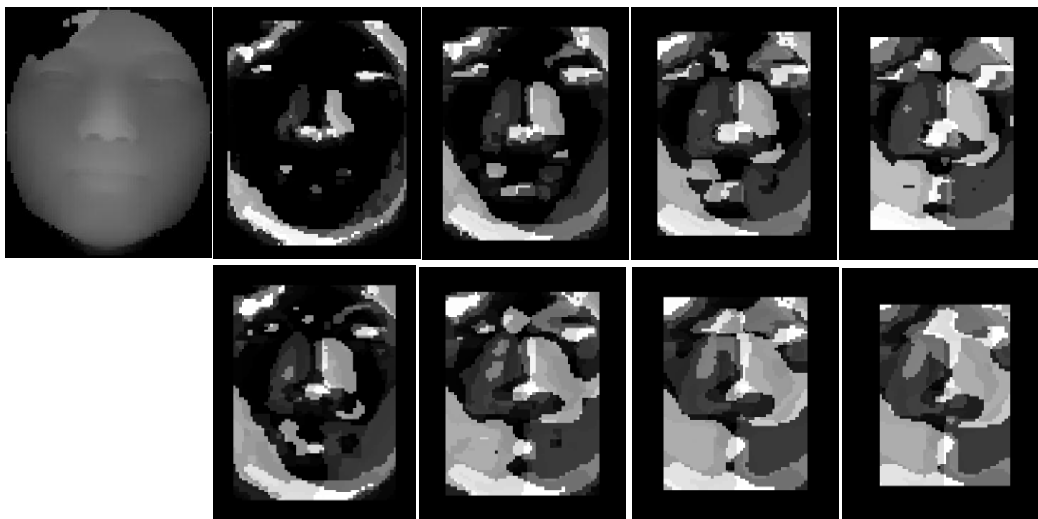


Figure 4.9 La représentation TPLBP d'une image de profondeur avec différentes valeurs de rayons et de tailles des patches w

4.4.3.2 Four-Patch Local Binary Patterns (FPLBP)

Le FPLBP est une autre variante de descripteur LBP basée sur les patches presque avec la même idée de TPLBP. Cependant, il y a une différence dans le nombre d'anneaux et des patches utilisés pour la comparaison. L'opérateur FPLBP utilise deux anneaux avec deux rayons $r1$ et $r2$ centrés dans un pixel de l'image. Les S patches sont uniformément répartis autour de ces deux anneaux (voir figure 4.10). Dans ce cas, la comparaison est produite entre deux centres de patches symétriques dans l'anneau intérieur et deux centres de patches symétriques dans l'anneau extérieur positionnés le long d'un cercle dans le sens des aiguilles d'une montre α ($\alpha = 1$) [121]. Après la comparaison, un bit dans chaque pixel de l'image est défini en fonction de celui des deux patches les plus similaires. Pour les S patches le long de chaque cercle, nous avons $S/2$ paires de centres symétriques qui sont la longueur des codes binaires produits. Le code FPLBP est donné par :

$$FPLBP = \sum_{i=0}^{\frac{S}{2}-1} f(d(P_{1,i}, P_{2,i+\alpha \bmod(S)}) - d(P_{1,i+\frac{S}{2}}, P_{2,i+\frac{S}{2}+\alpha \bmod(S)})) 2^i \quad (4.9)$$

L'exemple ci-dessous explique le codage de pixel selon le descripteur FPLBP avec $\alpha=2$, $S=8$, et $w=3$:

$$FPLBP = (f((P_{1,0}, P_{2,1}) - d(P_{1,4}, P_{2,5})) 2^0 + f((P_{1,1}, P_{2,2}) - d(P_{1,5}, P_{2,6})) 2^1 \\ + f((P_{1,2}, P_{2,3}) - d(P_{1,6}, P_{2,7})) 2^2 + f((P_{1,3}, P_{2,4}) - d(P_{1,7}, P_{2,8})) 2^3)$$

La figure 4.10 illustre une représentation de l'image de profondeur avec le descripteur FPLBP. La rangée du haut de gauche à droite : l'image de profondeur d'entrée, le code image FPLBP ($r1=1, r2=5, S=8$) avec $w = 3, 5, 7$ et 9 . La rangée du bas de gauche à droite : le code image FPLBP ($r1=1, r2=5, S=12$) avec $w=3, 5, 7$ et 9 .

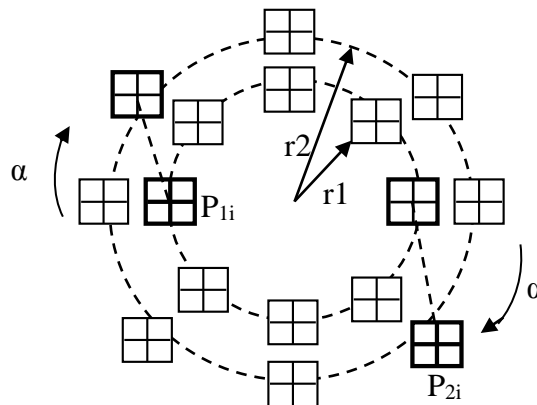


Figure 4.10 La procédure du descripteur FPLBP



Figure 4.11 La représentation FPLBP d'une image de profondeur avec différentes valeurs de tailles de patches w et de nombre des patches voisins S

4.4.5 Binarized Statistical Image Features (BSIF)

Le BSIF est un nouveau descripteur proposé par *J. Kannala et E. Rahtu* [127] en 2012. Le descripteur BSIF est inspiré par la méthodologie LBP et LPQ. Un code binaire pour chaque pixel est calculé en projetant linéairement des patches locaux de l'image sur un sous-espace, dont les vecteurs de base sont appris à partir d'images naturelles par l'analyse en composantes indépendantes (Independent Component Analysis, ICA), en utilisant la binarisation des coordonnées dans cette base par le seuillage [144]. L'objectif du descripteur BSIF est d'obtenir une représentation significative d'une image basée sur les caractéristiques statistiques. Le code binaire de l'image d'entrée est calculé pour chaque pixel en utilisant sa réponse à un ensemble fixe de filtres qui sont automatiquement appris basé sur des propriétés statistiques d'un petit ensemble d'images naturelles. L'ICA est utilisée pour l'apprentissage de l'ensemble des filtres linéaires en maximisant l'indépendance statistique des réponses (de filtres) [141]. Les filtres utilisés sont tirés d'un ensemble de 13 images naturelles donné par *A. Hyvärinen et al* dans [145] (voir figure 4.12). En supposant que X est notre image d'entrée et W_i le filtre linéaire de taille $l * l$ (voir la figure 3.13), la réponse du filtre S_i est donnée par l'équation suivante :

$$S_i = \sum_{u,v} W_i(u,v)X(u,v) = w^T x \quad (4.10)$$

Où w et x comprennent les pixels de W_i et X respectivement. Les n filtres utilisés sont assemblés dans une matrice W de taille $n * l^2$, le nombre de ligne n représente le nombre de filtre utilisé et le nombre de colonne l^2 représente les pixels du filtre sous la notation vecteur. Chaque bit dans le code BSIF final est associé à un filtre différent et la longueur de la chaîne de bits (n) détermine le nombre de filtres utilisés. La longueur de la chaîne de bits n avec la taille du filtre l sont des paramètres variables pour évaluer le descripteur BSIF. Toutes les

réponses sont calculés à la fois $s = W \cdot x$. La chaîne de bits (binaire) b est obtenue par la binarisation de chaque élément S_i . La caractéristique binarisée b est calculée par :

$$b_i = \begin{cases} 0 & \text{if } s_i > 0 \\ 1 & \text{ailleurs} \end{cases} \quad (4.11)$$

L'apprentissage du filtre linéaire avec l'algorithme ICA repose sur la maximisation de l'indépendance statistique des réponses de filtre. Cette approche offre des bonnes caractéristiques pour le traitement d'image [127, 145]. Pour cela, il faut décomposer la matrice de filtres W en deux parties : $S = Wx = UVx = Uz$ (4.12)

Où $Vx=z$ et U une matrice carrée ($n * n$) qui sera estimée par l'algorithme ICA. La matrice V effectue le prétraitement canonique, c.à.d, la réduction de la dimension des échantillons d'apprentissage x [145]. La réduction de données basée sur la PCA est utilisée pour réduire la taille de x , ne gardant que les premières composantes principales qui sont divisées par leur écart-type pour obtenir les échantillons de données. Finalement, on obtient la matrice du filtre $W = UV$. Le vecteur de sortie est binarisé par le seuillage de chacun de ses éléments à zéro. Enfin, nous obtenons une chaîne binaire traitée comme un mot de code caractérisant le contenu de la zone de voisinage local sur un certain endroit dans l'image d'entrée. La figure 3.14 illustre la représentation BSIF d'une image de profondeur avec différentes taille du filtre (l) et différentes longueur de la chaîne de bits n .

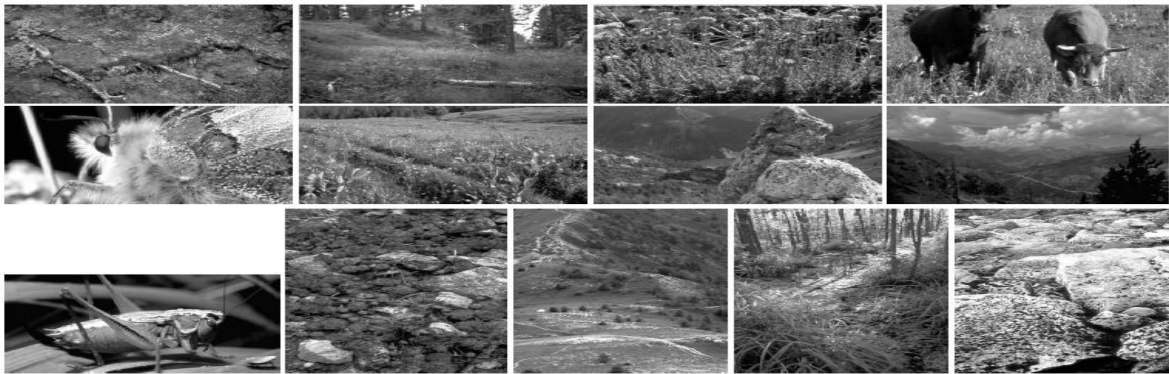


Figure 4.12 Les 13 images naturelles utilisées pour l'apprentissage des filtres dans le descripteur BSIF [97]

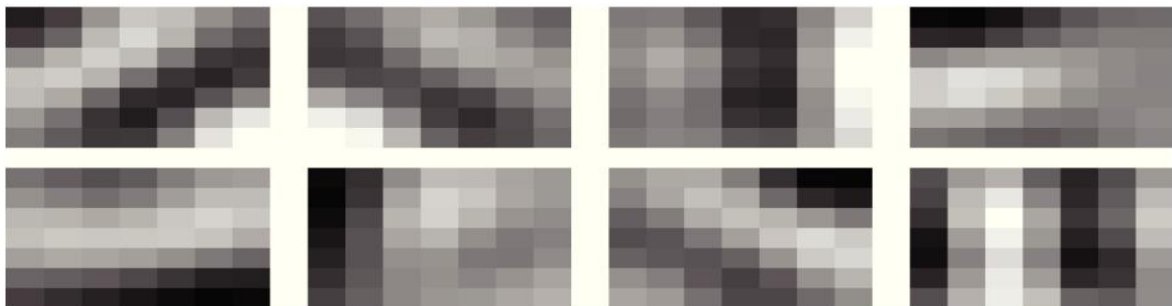


Figure 4.13 Filtres de taille 9×9 ($l = 9$)

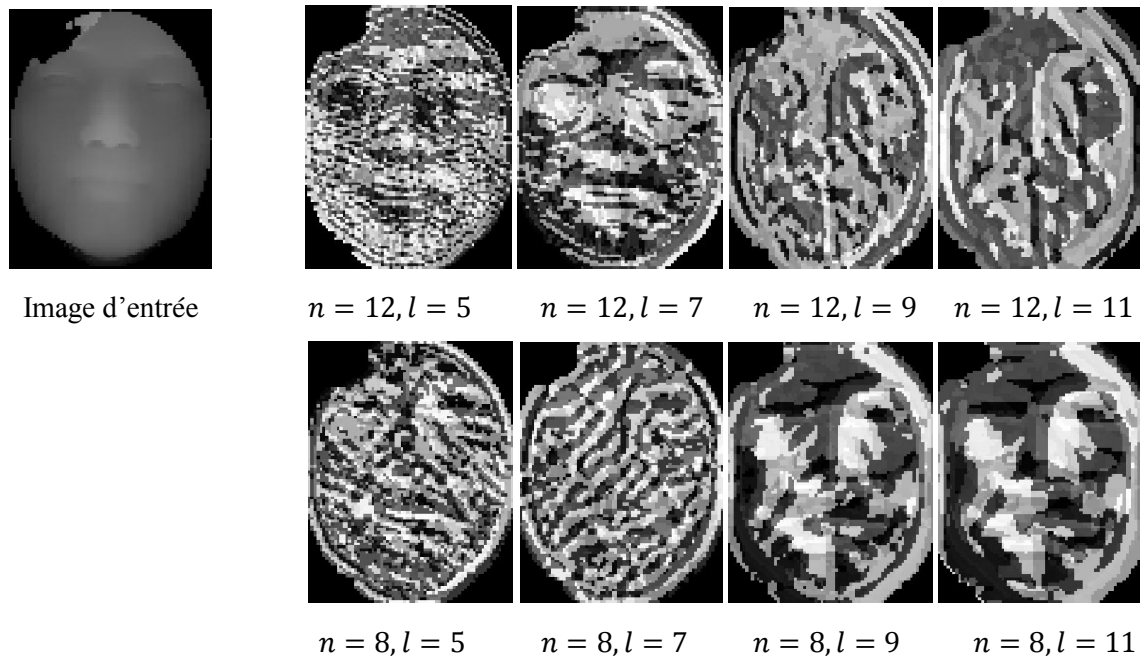


Figure 4.14 La représentation BSIF d'une image de profondeur avec différentes tailles (l) du filtre et différentes longueur de la chaîne de bits n

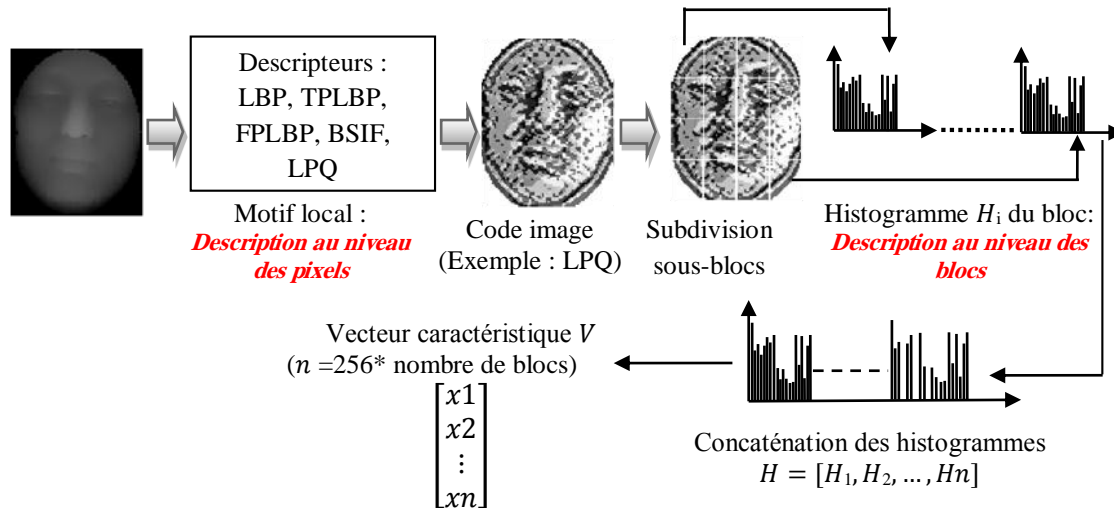
4.5 Extraction des caractéristiques d'histogrammes

Après le traitement des images faciales avec l'un des descripteurs locaux, nous obtenons des images étiquetées (le code image) avec le modèle correspondant. Ensuite, l'image de visage est subdivisée en différents blocs rectangulaires. L'histogramme de chaque bloc est calculé. Ensuite, tous les histogrammes sont concaténés dans un vecteur caractéristique unique. La figure 4.15 illustre notre méthode d'extraction des caractéristiques. L'histogramme représente les informations concernant les micros structures telles que les bords, les tâches et les zones plates dans les régions locales [146]. Par conséquent, les histogrammes sont utilisés comme des caractéristiques discriminantes pour représenter l'information locale de l'image faciale. L'histogramme (H) de l'image d'entrée (I) est calculé par :

$$H(i) = \sum_{x,y} B(I_p(x,y) = i), i \in [0, 2^p - 1], B(v) \begin{cases} 1 & \text{lorsque } v \text{ est vrai} \\ 0 & \text{autrement} \end{cases} \quad (4.13)$$

Où P est le nombre de bits. La taille du vecteur de caractéristiques de sortie est de 256 éléments multiplié par le nombre de blocs de chaque image et multiplié aussi par le nombre de bits de codage utilisé. Donc, un vecteur avec une grande dimension est produit pour la représentation de l'image faciale. La haute dimensionnalité des vecteurs caractéristiques conduit à un temps de calcul élevé ainsi qu'à une redondance de l'information entre les caractéristiques extraites, par conséquent, une faiblesse au niveau de la classification. Dans le but de limiter ces inconvénients, la réduction de dimensionnalité est effectuée. Cette étape est généralement

réalisée avant d'appliquer les algorithmes de classification pour les systèmes de reconnaissance de visage.



4.6 Réduction de dimension basée sur OLPP

Dans la pratique en reconnaissance de visage, il y a un problème largement connu, c'est la grande dimensionnalité des données d'entrée. Chercher à réduire la dimensionnalité d'un ensemble de données devient de plus en plus indispensable pour les applications de la biométrie faciale. Dans notre système de vérification dans ce chapitre, chaque image de visage est représentée par un vecteur de caractéristiques. La grande dimension de ces vecteurs provoque des dégradations à la fois, pour les performances de vérification et pour le temps de calcul. Dans cette section, nous présentons l'algorithme OLPP. Pour les algorithmes PCA et PCA+EFM, consulter le chapitre 3, section 3.4.

La méthode OLPP appelée aussi Laplacian Faces Orthogonale, est une méthode de reconnaissance de visage basée sur l'apparence proposée par *C. Deng et al* [128] en 2006. Les fonctions de base de Laplacian Face sont les vecteurs propres associés aux plus petites valeurs propres, l'OLPP requies ces fonctions de base pour être orthogonale et peut avoir plus de puissance de discrimination par rapport à l'algorithme Laplacian Face (LPP). Soit $\{x_1, x_2, \dots, x_k\}$ un ensemble de données dans R^l , l'OLPP cherche à trouver une matrice de transformation A qui projette les k points de l'ensemble président à un ensemble de points $\{y_1, y_2, \dots, y_k\}$ dans R^m lorsque $m \ll l$ et $y_i = A^T \cdot x_i$. L'algorithme OLPP comprend cinq étapes de la façon suivante :

- *Etape 1* : projection PCA, les images faciales x_i sont projetées dans le sous-espace PCA. La matrice de transformation PCA est représentée par W_{PCA} . Les caractéristiques non corrélées sont extraites. Le rang de la nouvelle matrice de données est égal au nombre de caractéristiques.
- *Etape 2* : Construire le graphique de contiguïté basé sur le k plus proches voisins, soit G un graphe dont les k nœuds. Le $i^{\text{ème}}$ nœud correspond aux images faciales x_i , Une arête est placé entre les nœuds i et j , x_i est parmi les p ($p \in \mathbb{N}$) plus proches voisins de x_j ou x_j est parmi les p plus proches voisins de x_i .
- *Etape 3* : Calculer les poids

$$W(i, j) = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{t}} & \text{Si les nœud } i \text{ et } j \text{ sont connectées} \\ 0 & \text{autrement} \end{cases} \quad (4.14)$$

Où t est une constante appropriée. La justification de ce poids et le réglage du paramètre t peut se référer à [147]. La matrice de poids W_{ij} des graphes G modélise la structure locale de la collection de visage.

- *Etape 4*: Les fonctions orthogonales de base

Soit D une matrice diagonale symétrique, dont les entrées sont la somme des colonnes ou des lignes de W_{ij} . $D_{ii} = \sum_j W_{ji}$, la matrice de Laplacian L est définie par : $L = D - W$. Soient $\{a_1, a_2, \dots, a_k\}$ les vecteurs de base orthogonales, nous définissons :

$$A^{(k-1)} = [a_1, a_2, \dots, a_{k-1}] \quad (4.15)$$

$$B^{(k-1)} = [A^{(k-1)}]^T (XDXT)^{-1} A^{(k-1)} \quad (4.16)$$

Où $X = [x_1, x_2, \dots, x_k]$. Ensuite, le procédé de réduction de dimensions peut être utilisé pour conserver les caractéristiques intrinsèques et rejeter les caractéristiques négligeables dans le sous-espace collecteur. Le vecteur de base orthogonale $[a_1, a_2, \dots, a_{k-1}]$ est calculé par le calcul de a_1 , le vecteur propre de $(XDXT)^{-1}XLX^T$ associé aux plus petites valeurs propres, et calculer aussi a_k , le vecteur propre de :

$$M^{(k)} = \left\{ I - (XDXT)^{-1} A^{(k-1)} [B^{(k-1)}]^{-1} [A^{(k-1)}]^T \right\} \cdot (XDXT)^{-1} XLX^T \quad (4.17)$$

associé avec les plus petites valeurs propres de $M^{(k)}$. Où $I = AA^T$ (A est une matrice orthogonale).

- *Etape 4* : La matrice de transformation OLPP

Soit $W_{OLPP} = [a_1, a_2, \dots, a_1]$, la transformation est donnée par :

$$\begin{aligned} x &\longrightarrow y = W^T x \\ W &= W_{PCA} W_{OLPP} \end{aligned} \quad (4.18)$$

Où y est la représentation de l-dimension de l'image du visage x , et W est la matrice de transformation finale.

4.7 Résultats expérimentaux et discussion

4.7.1 Base de données et protocole de vérification

Cette section donne des détails sur les résultats expérimentaux du système proposé. Pour démontrer l'efficacité de l'approche proposée. Les expériences ont été réalisées sur la base de données de visage CASIA3DV4 qui contient 123 personnes. Dans ce travail, nous utilisons 1.845 images où chaque personne est représentée par 15 modèles différents. La figure 4.17 montre les 15 images de profondeur de la même personne dans la BDD de visage CASIA3DV4. Comme le montre le tableau 4.1, notre protocole comprend trois partitions de données : Apprentissage, Evaluation et Test.

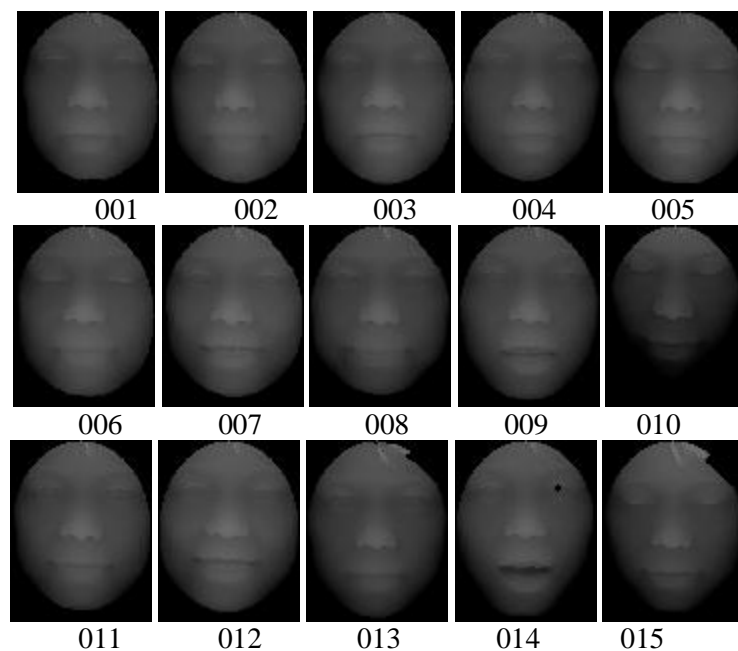


Figure 4.16 Image de profondeur de la même personne, première ligne variations d'illuminations, deuxième ligne variation d'expressions, troisième ligne variations d'expressions sous différentes illuminations

Tableau 4.1 Protocole utilisé pour la vérification

Dataset	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	500 images (3, 5, 11, 12, 13)	150 images (1:15)

4.7.2 Analyse de résultats et discussion

Dans les premières expériences, le système de vérification 3D du visage proposé est évalué avec différents descripteurs locaux : LBP, TPLBP, FPLBP, BSIF et LPQ avec trois méthodes de réduction de dimension, PCA, OLPP et PCA + EFM.

Dans la deuxième expérience, la combinaison des caractéristiques du descripteur LPQ avec tous les autres descripteurs est utilisée avec la meilleure méthode de réduction de dimension.

Les résultats expérimentaux ont été obtenus avec les paramètres optimaux de chaque descripteur. Le descripteur LPQ est basé sur la quantification de la transformée de Fourier de la phase dans les voisins d'une fenêtre de taille 5*5 à chaque pixel. La taille optimal du filtre linéaire utilisé pour le descripteur BSIF est 17*17. Tous les paramètres sont choisis suite aux résultats des expériences. La performance de notre système est testée avec le descripteur LBP. Le code d'image avec $p=16$ et $r=2$ nous donne le meilleur résultat. Les paramètres optimaux des descripteurs sont choisis expérimentalement. Ces paramètres sont montrés dans le tableau 4.2.

Tableau 4.2 Paramètres optimaux des descripteurs utilisés

Descripteur	Paramètres optimaux
LPB	$p=16, r=2$
BSIF	$l=17$
LPQ	$R=5$
TPLBP	$r1=3 \quad w=8 \quad s=8 \quad \tau=0.01 \quad r1=1$
FPLBP	$r2=5 \quad w=12 \quad \tau=0.01$

La figure 4.17 illustre une image de profondeur et le code de l'image LBP, TPLBP, FPLBP, BSIF et LPQ avec les paramètres optimaux.

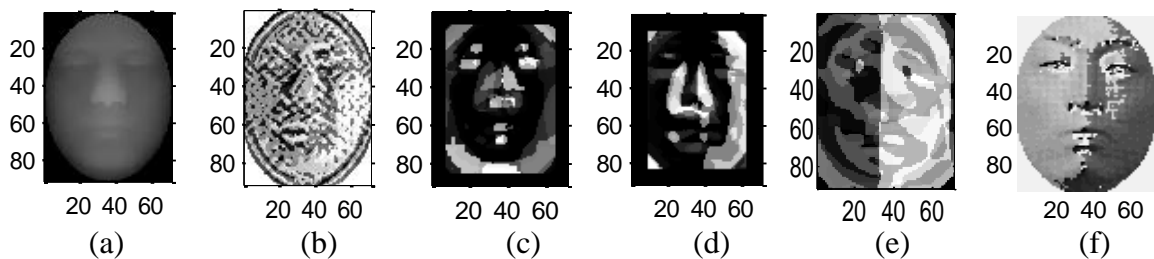


Figure 4.17 Représentation du visage basée sur les descripteurs locaux; (a) image d'entrée; (b) code image LPQ; (c) code image TPLBP, (d) code image FPLBP, (e) code image BSIF, (f) code image LBP

Les tableaux 4. 3, 4.4 et 4.5 montrent les performances de la vérification avec les méthodes, PCA, OLPP et PCA + EFM avec les cinq descripteurs locaux.

Tableau 4.3 Performances de la vérification avec PCA durant l'évaluation et le test

Descripteur local	Evaluation	Test			
	TEE (%)	TFA (%)	TFR (%)	HTER (%)	TV (%)
TPLBP	10.77	8.38	8.40	8.39	83.22
FPLBP	8.74	6.54	6.80	6.67	86.65
BSIF	8.79	7.25	7.05	7.15	85.66
LPQ	7.04	7.13	5.40	6.62	87.46

Tableau 4.4 Performances de la vérification avec OLPP durant l'évaluation et le test

Descripteur local	Evaluation	Test			
	TEE (%)	TFA (%)	TFR (%)	HTER (%)	TV (%)
TPLBP	5.15	2.52	4.60	3.56	92.88
FPLBP	4.78	1.62	4.40	3.01	93.98
BSIF	3.60	2.10	3.70	2.90	94.19
LPQ	2.35	1.17	2.80	1.98	96.02

Tableau 4.5 Performances de la vérification avec PCA+EFM durant l'évaluation et le test

Descripteur local	Evaluation	Test				
	TEE (%)	TFA (%)	TFR (%)	HTER (%)	TV (%)	TC (s)
LBP	3.03	0.91	3.01	1.96	96.08	1.27
TPLBP	1.90	0.71	2.80	1.75	96.48	1.13
FPLBP	1.90	0.71	2.60	1.65	96.68	1.91
BSIF	2.42	1.19	1.55	1.37	97.26	1.18
LPQ	1.20	0.21	1.60	0.90	98.18	1.48

Nous pouvons constater que les deux algorithmes OLPP et PCA+ EFM sont parvenus à une amélioration considérable du taux de vérification par rapport à l'algorithme de base PCA avec plus de 9% à 11%. Par exemple avec l'opérateur TPLBP, le taux de vérification est égal à 83.22%, 92.88% et 96.48% avec PCA, OLPP et PCA+EFM respectivement. Aussi, en utilisant le descripteur FPLBP nous obtenons 86.65%, 93.98% et 96.68% en utilisant PCA, OLPP et PCA+EFM respectivement. Les résultats présentés dans les tableaux 4.3, 4.4 et 4.5 au cours d'évaluations et des tests indiquent que LPQ et BSIF conduisent aux hautes performances de vérification ; nous obtenons un taux de vérification de 97,26% et 98,18% pour BSIF et LPQ et respectivement. D'autre part, nous pouvons confirmer ces résultats à travers les courbes ROCs des différentes expériences qui sont illustrées dans les figures 4.18, 4.19 et 4.20. La courbe ROC trace le taux de vérification en fonction de taux de fausse acceptation. La BSIF arrive à la deuxième meilleure performance de vérification après le LPQ avec toutes les méthodes de réduction de la dimension, sauf dans un cas lorsque nous utilisons l'algorithme de base PCA.

Nous obtenons un taux de vérification faible égal à 85,66% comme il est illustré dans le tableau 4.3. Les histogrammes LPQ explorent des caractéristiques très discriminantes pour la vérification, et ils surpassent toutes les autres caractéristiques des différents descripteurs avec un taux de vérification égal à 98.18% et un TEE égal à 1.20 %.

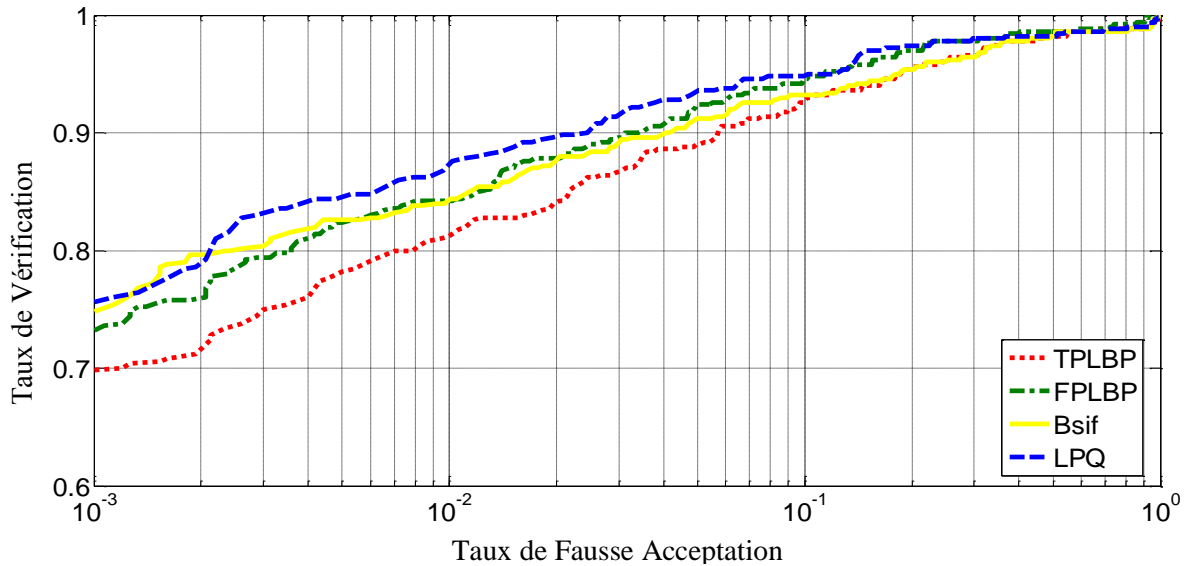


Figure 4.18 Courbes ROCs avec PCA

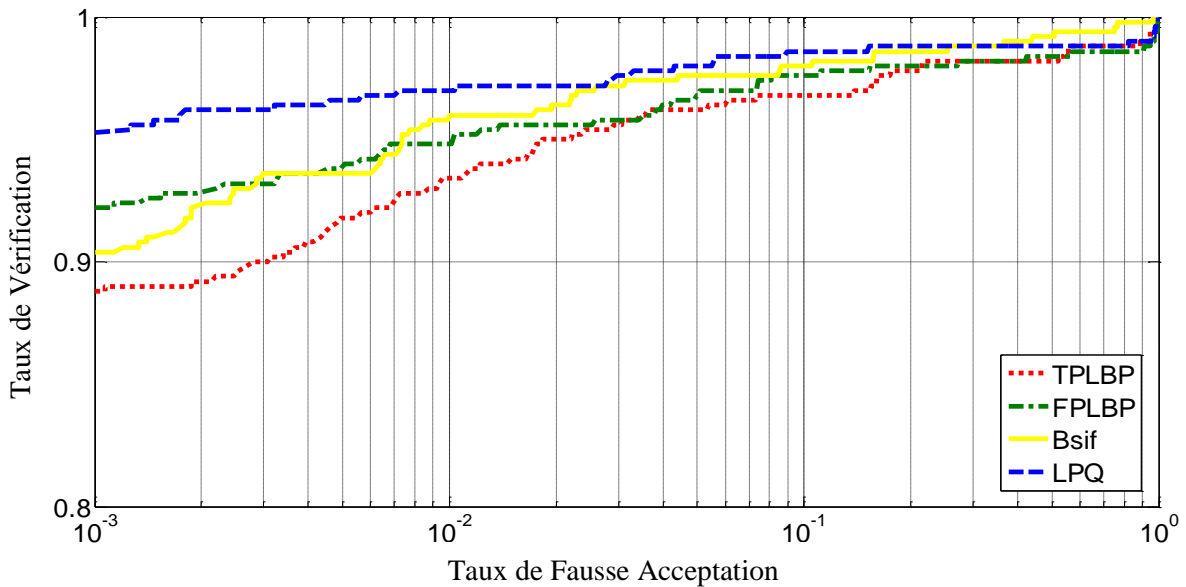


Figure 4.19 Courbes ROCs avec OLPP

Nous pouvons dire que le descripteur LPQ basé sur PCA + EFM est l'un des meilleurs cadres pour aborder l'effet de l'illumination et des d'expressions faciales dans le système de vérification du visage. Nous attribuons ce résultat au fait que les illuminations et les variations des expressions faciales n'ont aucun effet sur le domaine fréquentiel qui est représenté par le descripteur LPQ. Basé sur ces résultats, dans les deuxièmes expériences, la combinaison du descripteur LPQ avec tous les descripteurs est utilisée.

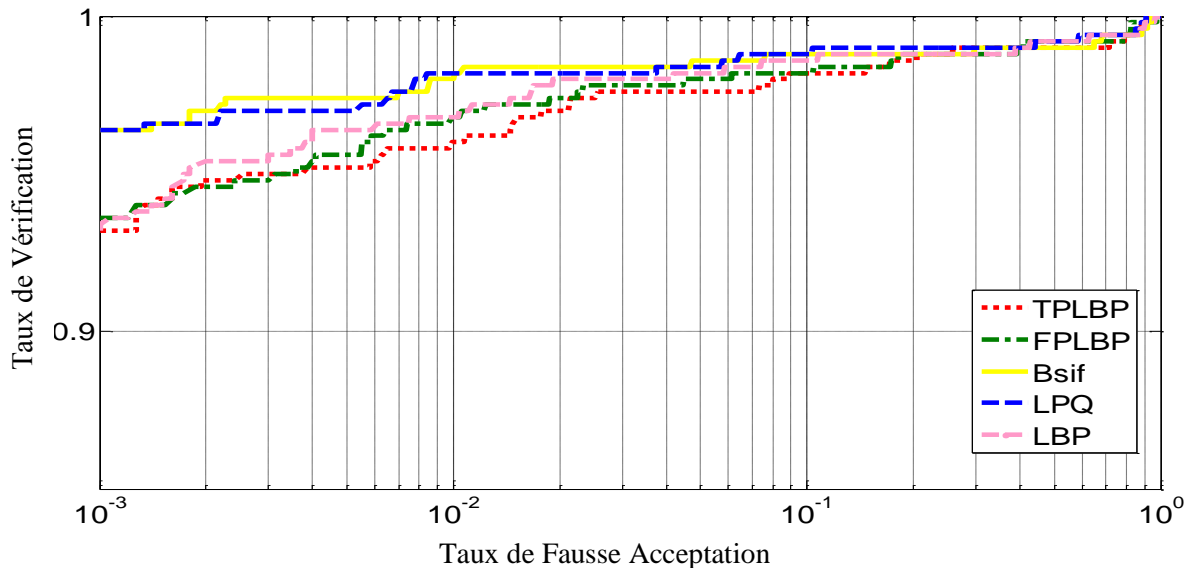


Figure 4.20 Courbes ROCs avec PCA+EFM

Le tableau 4.6 présente les performances de vérification de la combinaison LPQ avec les autres descripteurs basés sur PCA+EFM.

Tableau 4.6 Performances de la vérification utilisant PCA+EFM basé sur la combinaison de LPQ avec LBP, FPLBP et BSIF

Descripteur Local	Evaluation	Test				
	TEE (%)	TFA (%)	TFR (%)	HTER (%)	TV (%)	TC (s)
LPQ_LBP	1.60	0.52	2.10	1.31	97.37	2.40
LPQ_TPLBP	1.80	0.44	2.06	1.25	97.49	2.80
LPQ_BSIF	1.40	0.33	1.93	1.13	97.73	3.03
LPQ_FPLBP	1.30	0.19	1.57	0.88	98.22	2.99

Dans ce tableau 4.6 les performances de vérification au cours de l'ensemble de test en utilisant LPQ_FPLBP (TV= 98,22%, HTER = 0,88%) surpasse à la fois LPQ_LBP, LPQ_TPLBP et LPQ_BSIF. Nous pouvons dire que la combinaison des histogrammes des caractéristiques locales LPQ_FPLBP basée sur la PCA+EFM lorsqu'on utilise des images de profondeur 3D donne des caractéristiques plus distinctes. Cette approche optimale produit une amélioration significative de la performance du système de vérification visage 3D.

Les résultats représentés dans le tableau 4.5 et 4.6 montrent que les méthodes étendues TPLBP et FPLBP peuvent surpasser le descripteur LBP, mais il n'y a pas une grande différence entre eux. Nous pouvons voir qu'il n'y a pas de gain de performance significatif dans les méthodes combinées. Le descripteur de LPQ dépasse certaines méthodes combinées telles que : LPQ_TPLBP, LPQ_BSIF et LPQ_LBP. Cela est dû au fait que ces descripteurs partagent

beaucoup de caractéristiques similaires (redondance) qui ont conduit à un affaiblissement du processus de classification.

Toutes nos expériences sont implémentées par le MATLAB R2010b sur un PC avec 2.53 GHz Intel Core i5 CPU et une RAM de 4 Go. Dans les tableaux 4.5 et 4.6, nous présentons le temps de calcul (TC) de chaque méthode. Il n'y a pas de TC significative entre toutes les expériences (1,90 (s) à 1,13 (s)), car dans toutes les méthodes proposées, nous utilisons presque la même taille de vecteurs caractéristiques. Comme nous pouvons l'observer dans le tableau 4.6, le TC des méthodes combinées a augmenté. Ceci s'explique dans ce cas par la concaténation de deux vecteurs caractéristiques où le premier est le descripteur LPQ et le second est l'un des autres descripteurs locaux utilisés.

Nous examinons aussi les effets de la variation de la taille des vecteurs caractéristiques. Nous pouvons déterminer les meilleures dimensions du vecteur caractéristique expérimentalement en variant la taille entre 10 et 100. La figure 4.21 illustre la variation des taux de vérification à travers la dimension du vecteur caractéristique.

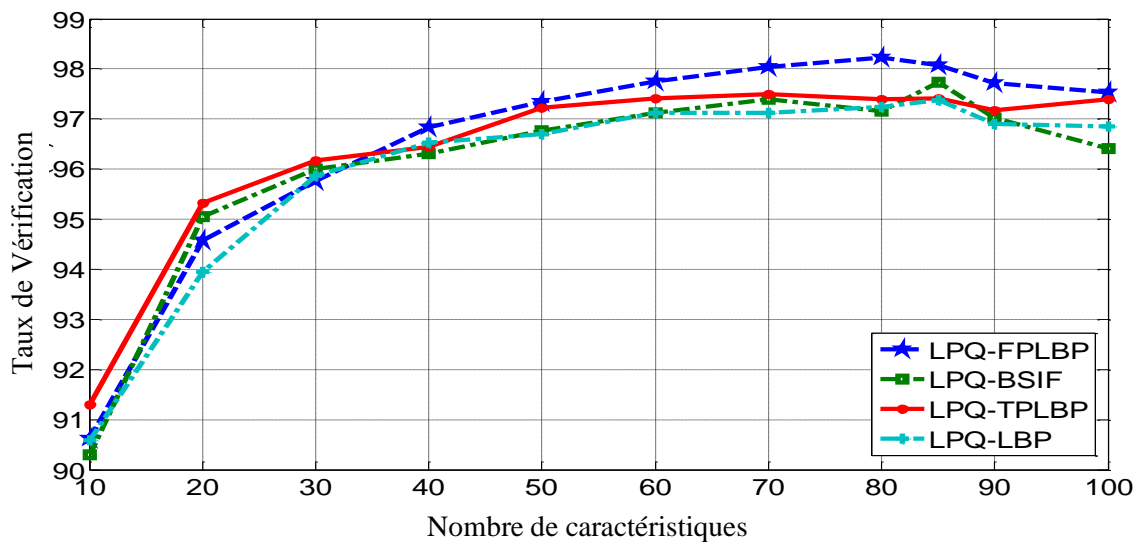


Figure 4.21 Variations de taux de vérification en fonction du nombre de caractéristiques

À partir de cette figure nous affirmons que le vecteur de projection discriminante optimale des quatre méthodes combinées (LPQ_LBP, LPQ_FPLBP, LPQ_BSIF et LPQ_TPLBP) en utilisant notre méthode PCA + EFM est situé entre 70 et 90.

Le tableau 4.7 présente une comparaison de nos meilleurs résultats en termes de taux de vérification avec plusieurs techniques existantes dans la littérature.

Tableau 4.7 Comparaison des taux de vérification avec l'état de l'art

Auteurs	Méthodes	Base de données	TV (%)
<i>X. Wang</i> [148]	ICP, PCA, Gabor, LBP, Corresponding Point Direction Measure (CPDM)	CASIA3DV4	91.71
<i>Y. A. Li</i> [149]	Caractéristique géométrique, ICP, LDA, géodésique.	CASIA3DV4	91.10
<i>Y.Ming</i> [10]	Régression spectrale orthogonale, PCA courbure, métrique, ICP, NN	CASIA3DV4	96.25
<i>Yue Ming</i> [11]	Gabor, Régression Spectrale Patchée, LDA	CASIA3DV4	96.35
<i>Ouamane, A</i> [150]	LBP, SIFT, SLF, SVM	CASIA3DV4	96.26
<i>Notre méthode</i>	LPQ_FPLBP, PCA+EFM, SVM	CASIA3DV4	98.22

4.8 Conclusion

Dans ce chapitre, nous avons présenté une validation expérimentale de notre système : la vérification automatique du visage 3D en présence des variations d'illuminations et d'expressions faciales en utilisant des images de profondeur. Cinq des meilleurs descripteurs locaux LBP, LPQ, BSIF, TPLBP et FPLBP ont été utilisés. Les histogrammes de chaque descripteur sont extraits sur des blocs rectangulaires dans l'image faciale et concaténés dans un seul vecteur caractéristique. Afin de réduire la grande dimensionnalité de ces vecteurs, trois méthodes de réduction de dimension sont évaluées et comparées : PCA, PCA+EFM et OLPP. D'après nos résultats, nous avons démontré que la méthode combinée LPQ_FPLBP basée sur PCA + EFM donne des performances de vérification très élevées avec un taux de vérification de 98,22% et HTER de 0,88% durant l'ensemble de test. Nos résultats montrent que la méthode PCA + EFM a atteint une différence dans le TV pouvant aller jusqu'à 10%, et de 5,8% pour le TEE par rapport à la méthode de base PCA.

Chapitre 5

Vérification de V3D Poses /Angles d'Euler et Tenseurs

5.1 Introduction	86
5.2 Travaux de recherche autour de la vérification V3D poses	86
5.3 Vérification du visage basée sur les angles d'Euler	89
5.3.1 Aperçu de l'approche proposée	89
5.3.2 Génération de différentes poses avec la transformation d'Euler	90
5.3.3 Résultats et discussion	92
5.4 Vérification 3D de visage basée sur les tenseurs	95
5.4.1 Propriétés des tenseurs et modélisation multidimensionnelle	97
5.4.1.1 Définition d'un tenseur	97
5.4.1.2 Déploiement des tenseurs.....	97
5.4.1.3 Produit scalaire tensoriel.....	98
5.4.1.4 Produit extérieur.....	99
5.4.1.5 Opérateur n -mode produit	99
5.4.2 Décomposition tensorielle par HOSVD	99
5.4.3 Description de l'approche proposée	101
5.4.4 Résultats et discussion	103
5.5 Conclusion.....	107

5.1 Introduction

L'estimation de la pose de la tête est un processus clé pour l'analyse du comportement humain et pour les Interactions Homme Machine (IHM) [151]. Pour cette raison, de nombreuses applications pourraient bénéficier des systèmes automatiques d'estimation de la pose de la tête. La variation de poses a été identifiée comme l'une des problèmes importants non résolus et il gagne un grand intérêt dans la communauté de recherche en vision par ordinateur et en reconnaissance de formes. Ce chapitre est divisé en deux parties essentielles :

- La vérification 3D du visage en utilisant l'apprentissage par les angles d'Euler.
- La modélisation tensorielle basée sur la méthode proposée HOSVD+PCA+EFM en utilisant l'algèbre multilinéaire pour résoudre le problème de la variation de poses dans le système de vérification 3D du visage.

Dans un premier temps, nous introduisons une nouvelle méthode de vérification 3D de visage, robuste et efficace aux variations de poses basée sur l'apprentissage en utilisant les angles d'Euler ($\pm 60^\circ$ selon l'axe Y). En se basant sur les résultats obtenus dans le chapitre 4, nous utilisons les histogrammes du descripteur LPQ des images de profondeur pour la représentation des images faciales. L'avantage de notre approche est que nous pouvons évaluer un système de vérification de visage 3D en présence de poses même dans les bases de données des visages frontales en générant des variations de poses grâce à la transformation d'Euler. Dans la deuxième partie, nous proposons une mise en point d'une nouvelle approche pour la vérification de visage avec une grande variété de poses basée sur l'utilisation de l'algèbre multilinéaire et les tenseurs. Les défis illuminations, expressions et poses sont considérés comme des problèmes d'algèbre multilinéaire dans laquelle l'ensemble des images de profondeur sont représentées sous forme d'un tenseur (multilinéaire) d'ordre supérieur. Ce dernier est décomposé en se basant sur la méthode proposée HOSVD+ PCA + EFM afin de réduire sa grande dimension. Notre contribution à ce stade de la thèse est la généralisation multilinéaire de la méthode PCA+EFM.

5.2 Travaux de recherche autour de la vérification V3D poses

La possibilité d'éliminer l'influence des variations de poses, ainsi que l'invariabilité dans des conditions d'éclairage rend la reconnaissance faciale 3D plus efficace et robuste dans le monde réel. De nombreux algorithmes ont été déjà établis en termes de reconnaissance de visage frontale, lorsque la personne à reconnaître regarde directement le dispositif d'acquisition. Plus récemment, des méthodes de reconnaissance faciale non frontale ont été

proposées. La plupart de ces méthodes nécessitent un processus d'estimation des poses de la tête dans un espace tridimensionnel, dans lequel les angles de rotation sont déterminés automatiquement et la pose de la tête est corrigée pour éliminer les disparités de reconnaissance en raison de ces variations. Comme c'est indiqué dans un état de l'art récent [30], les protocoles pour les tests des systèmes de reconnaissance faciale en présence de poses ne sont pas encore unifiés, cela signifie que nous avons encore un long chemin pour construire un système de reconnaissance de visage robuste aux variations de poses. Il existe diverses approches qui abordent ce problème de la variation de poses en utilisant des méthodes 2D ainsi que des méthodes 3D. Nous résumons ci-dessous un ensemble intéressant de ces approches.

Georgios Passalis et al [152] ont utilisé la sélection des points de repère (Landmark) manuellement pour estimer la pose du visage et détecter les zones occluses. Un modèle de visage est enregistré et monté sur le scan. Lors du montage, la symétrie du visage est utilisée pour relever les défis de données manquantes (voir la figure 5.1). Les auteurs effectuent des comparaisons entre les scans interposés à l'aide d'une signature biométrique basée sur les ondelettes. Une limitation de cette approche est que l'extraction des points de repère lors de l'apprentissage se fait manuellement. Une solution pour résoudre ce types de problèmes est introduite par *Dibeklioglu et al* [153], les auteurs ont proposé une méthode pour détecter les points de repère automatiquement sur les images faciale de profondeur 3D à l'aide des courbures basées sur l'analyse heuristique pour permettre de corriger les poses dans un système qui permet d'identifier les personnes sous différentes variations de poses. Cependant, une limitation de ce procédé est qu'il est appliqué pour des images du visage avec des rotations selon l'axe Y inférieures à $\pm 45^\circ$.

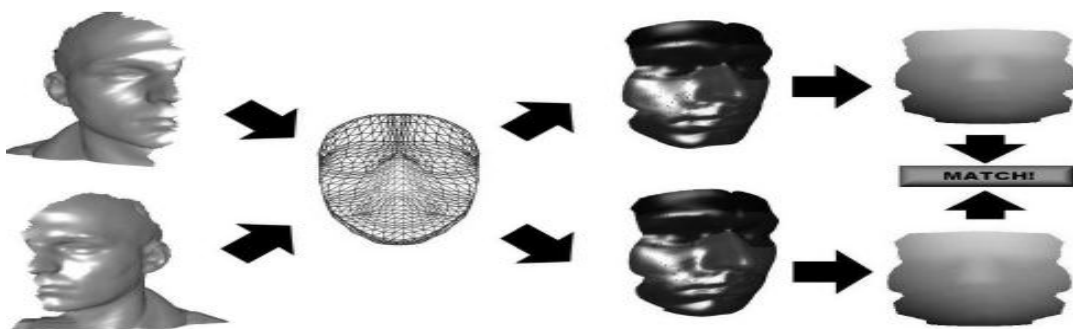


Figure 5.1 Résolution du problème de données manquantes utilisant la symétrie du visage sur les zones occultées [152]

Blanz et al [109] ont proposé une méthode holistique basée sur la reconstruction du visage en adaptant le modèle 3DMM sur des scans du visage 3D. Le 3DMM est construit en utilisant l'analyse en composantes principales sur des formes de visage 3D et des textures qui sont

obtenues à partir d'un scanner laser. Le 3DMM est considéré comme une approche de base pour traiter la variation de pose (cité plus 1625 fois d'après Google Scholar). Toutefois, les tests de reconnaissance de visage sont réalisés avec des visages sous des variations de pose qui ne dépassent pas $\pm 40^\circ$ selon l'axe Y. Dans [38], un algorithme a été proposé pour la correction automatique de la pose du visage. La pose d'un visage 3D avec sa texture est automatiquement corrigée en se basant sur le bout du nez (le processus de la détection est détaillé dans la section 2.3.2 du chapitre 2) en utilisant la transformation de Hotelling qui est basée sur l'algorithme PCA, dans lequel la matrice de rotation V aligne le nuage de points P sur ces axes principaux. Le ré-échantillonnage des visages sur une grille carrée uniforme a été utilisé pour assurer que tous les visages finissent avec la même résolution. Ceci est très important pour la précision de l'algorithme de correspondance qui est basée sur la mesure de distance point à point. Le filtre médian est utilisé sur la composante z du visage 3D (la profondeur) pour la suppression de flou. Cette méthode est évaluée aux données faciales avec de petites variations de poses inférieures à $\pm 15^\circ$ selon l'axe Y. La figure 5.2 montre le diagramme synoptique de l'algorithme proposé.

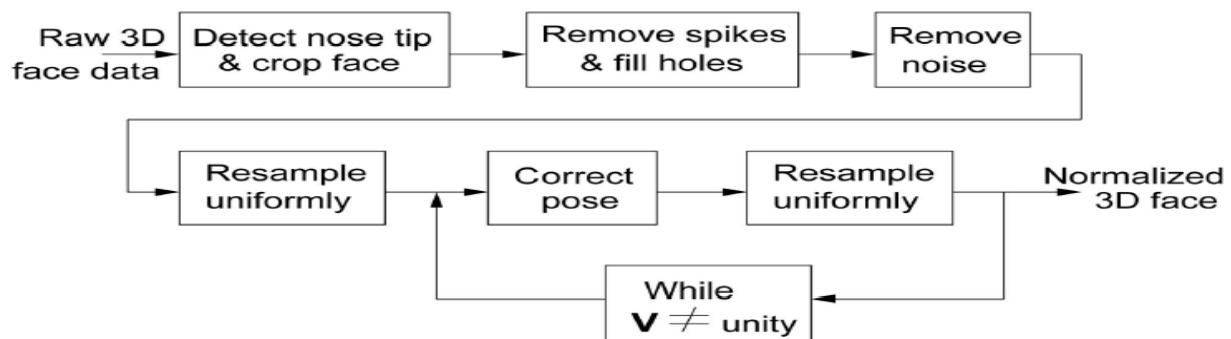


Figure 5.2 Normalisation de visage 3D proposée par *Mian et al* [38]

Une autre idée intéressante est présentée dans [154] nommée la signatures du profil tourné (Rotated Profile Signatures, RPS). La RPS est utilisée pour détecter le nez ainsi que la catégorisation des différentes poses. Le visage 3D est tourné de 0° à 180° avec un pas de 5° . À chaque étape, les points de "profil" les plus à droite sont extraites (voir figure 5.3) et adaptés à une grande variété de modèles de nez, le résultat est un score de similarité. Enfin, le score minimum indique l'emplacement correct du nez. Le point faible de cette méthode est qu'elle nécessite une grande variété des modèles de nez pour comparer et catégoriser les différents types de profils.

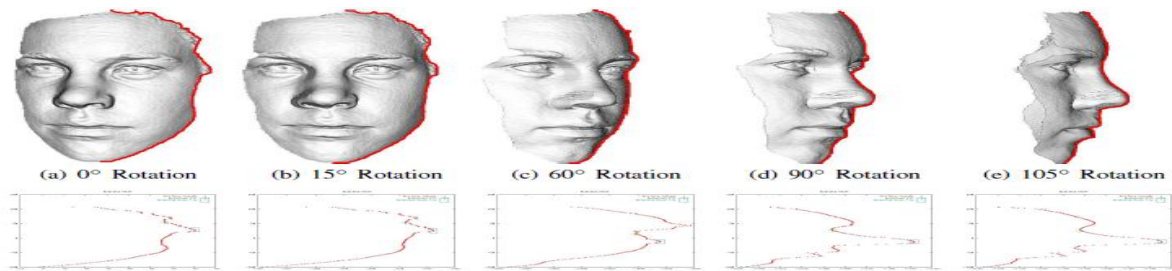


Figure 5.3 Signatures de profil tourné selon Y; les images du haut montrent les modèles 3D avec le profil extrait et les images du bas montrent les profils extraits associés [154]

Une autre méthode similaire proposée par *Rajwade et al* [155], les auteurs ont démontré une méthode de détection et de correction automatique de poses en utilisant la régression à support vecteurs (Support Vector Regression, SVR) sur les sous-bandes d'ondelettes. Cette technique est capable de classer les sujets de 3D pose de manière invariante avec une précision de $\pm 9^\circ$ dans les axes X et Y. *Wuming et al* [156] ont proposé un système de reconnaissance faciale 3D qui peut réduire l'influence causée par les variations de poses. La Forest de régression aléatoire (Random Regression Forest, RRF) est utilisée pour estimer les poses de visages. Puis, les modèles de visage 3D sont tournés dans la galerie comme ceux de la pose de test afin de générer un exemple de galerie spécifique pour l'appariement basée sur LBP. La figure 5.4 illustre la méthode RRF pour l'estimation de la pose.

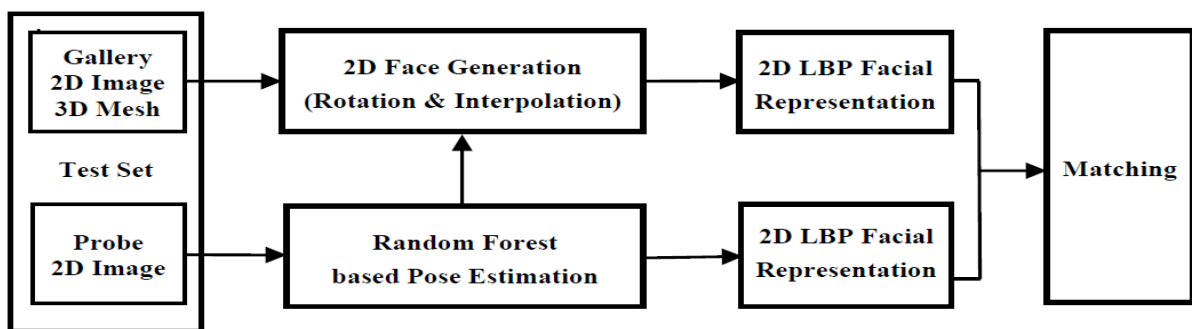


Figure 5.4 Estimation de la pose basée sur RRF [156]

5.3 Vérification du visage basée sur les angles d'Euler

5.3.1 Aperçu de l'approche proposée

Généralement, les approches proposées dans la littérature concernant la vérification de visage en présence des variations des poses sont basées sur l'étape de la correction de pose de la tête soit par des méthodes manuelles dans laquelle l'angle de rotation est mis à zéro degrés ou bien automatiquement avec l'estimation de l'angle d'orientation. En effet, ces processus généralement exigent un calcul mathématique complexe concernant les coordonnées du visage dans l'espace tridimensionnelle. D'autre part, ces approches nécessitent des bases de données

qui contiennent de différentes variations de poses pour chaque individu. Notre système de vérification de visage 3D proposé ne nécessite pas une étape de correction manuelle ou l'estimation d'angles d'orientation. Tout d'abord, la zone de visage est détectée avec les courbes IPC (voir chapitre 3). Ensuite, nous utilisons un visage avec la pose frontale et une expression neutre de chaque personne dans la base de données pour générer différentes poses de -60° jusqu'à $+60^\circ$ (voir la figure 5.6) en utilisant la transformation d'Euler. Après cela, sur la base de nos meilleurs résultats obtenus dans le chapitre 4, nous utilisons les histogrammes de descripteur LPQ multi-blocs pour la représentation de chaque image de profondeur. Ensuite, PCA+EFM est utilisée pour la réduction de dimension. Enfin, le processus de classification est basé le classificateur SVM multi classes pour faire la décision. L'organigramme de notre système proposé est illustré dans la figure 5.5.

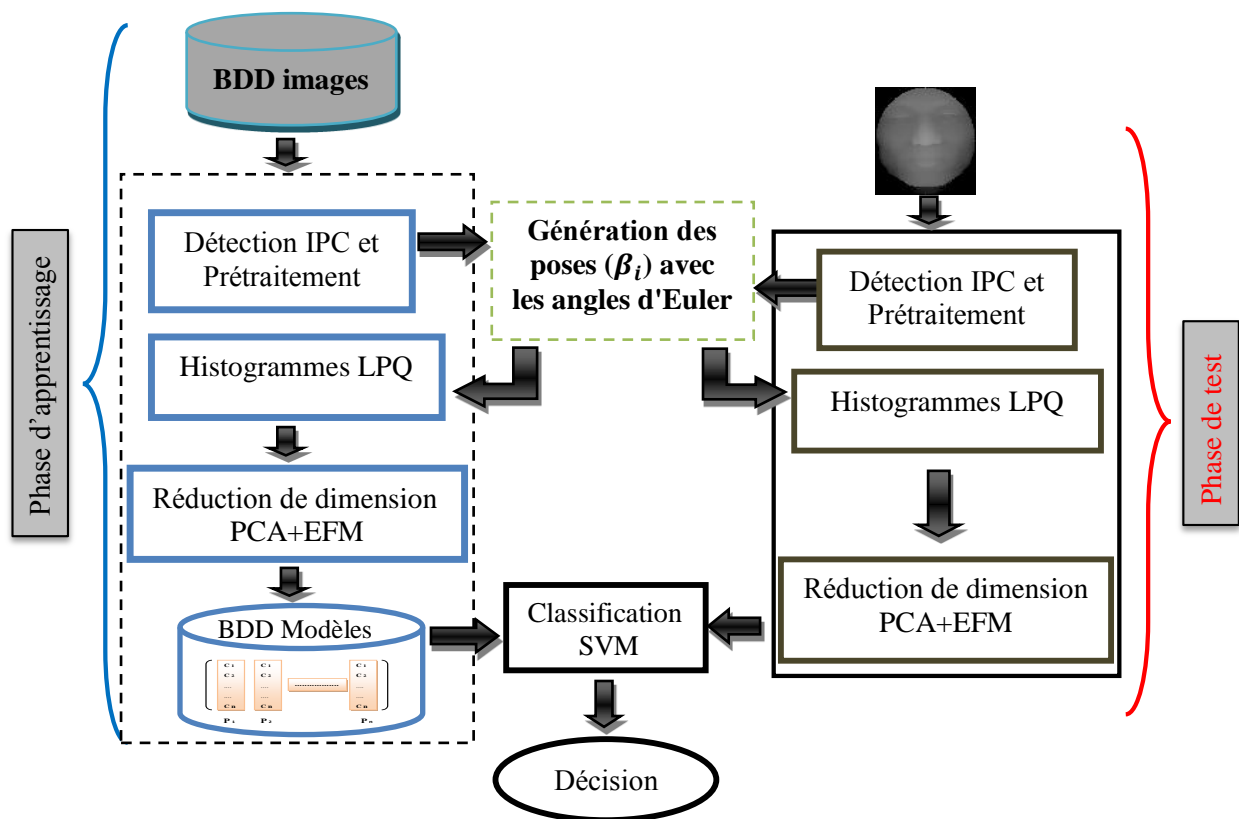


Figure 5.5 Organigramme de la méthode de vérification de visage 3D proposée

5.3.2 Génération de différentes poses avec la transformation d'Euler

La rotation de visage en 3D ou bien les angles d'Euler en trois dimensions sont donnés par une matrice de rotation 3D (3D Rotation Matrix, 3DRM). Cette dernière, est une matrice orthogonale de déterminant 1, ce qui peut s'exprimer par les équations suivantes :

$$R_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \quad (5.1)$$

$$R_y = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \quad (5.2)$$

$$R_z = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (5.3)$$

Où R_x, R_y, R_z sont des matrices de rotation selon les axes X, Y, Z respectivement. Nous considérons ainsi :

- La rotation de visage autour de l'axe X (angles Roll) est obtenue en variant l'angle α .
- La rotation de visage autour de l'axe Y (angles Yaw) est obtenue en variant l'angle β .
- La rotation de visage autour de l'axe Z (angles Pitch) est obtenue en variant l'angle γ .

La 3DRM dans l'espace 3D (X, Y, Z) est donnée par :

$$3DMR = (R_x * R_y * R_z) \quad (5.4)$$

Dans notre système, nous tournons les images de visage en 3D autour de l'axe Y selon les angles β . En mettant α et γ égal à 0° et nous calculons la 3DMR, nous obtenons :

$$3DMR = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} * \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} * \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$3DMR = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix}$$

Ensuite, le visage tourné en 3D est donné par :

$$\begin{pmatrix} x_i^\beta \\ y_i^\beta \\ z_i^\beta \end{pmatrix} = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} \quad (5.5)$$

Où un point dans le nuage de points d'entrée est défini par (x_i, y_i, z_i) ($i = 1, 2, \dots, N$) à partir d'une image de visage 3D qui se compose de N points, son point de sortie correspondant est défini par $(x_i^\beta, y_i^\beta, z_i^\beta)$. L'angle de la tête β_i ($i=1, 2, 3, \dots, 10$) varie selon ($\beta_i = -60^\circ, -50^\circ, -40^\circ, -30^\circ, -20^\circ, +20^\circ, +30^\circ, +40^\circ, +50^\circ, +60^\circ$). La figure 5.6 donne une illustration des différents angles de rotation de la tête β_i générés par la transformation d'Euler pour une image de profondeur.

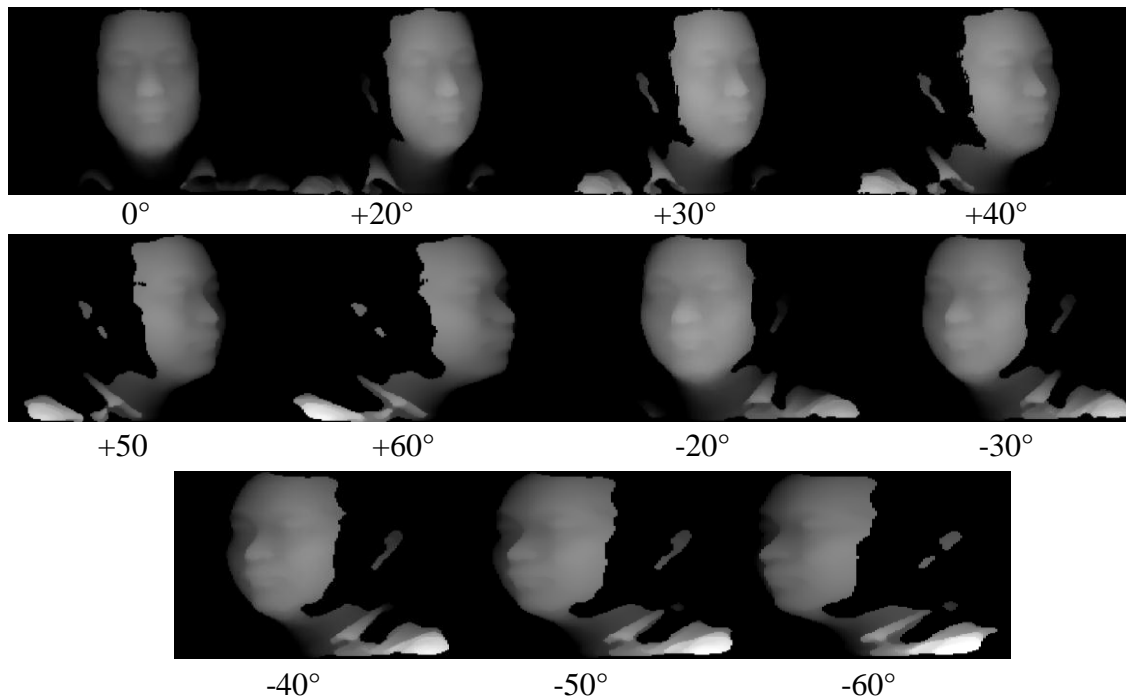


Figure 5.6 Les angles de rotation de la tête β_i générés par la transformation d'Euler pour une image de profondeur

5.3.3 Résultats et discussion

Afin de produire des résultats pour un autre scénario et d'évaluer l'effet de la présence d'une grande variation de poses sur notre système déjà présenté dans le chapitre 4, où nous avons évalués le système en présence des variations d'illuminations et d'expressions faciales, dans ce cas, des résultats excellents ont été obtenus avec un TV=98.18% et un HTER= 0.9 %. Dans cette section, notre scénario est basé sur le protocole représenté dans le tableau 5.1. Le même protocole utilisé dans le chapitre 4 est suivi (tableau 4.1) nous ajoutons différentes poses de la tête illustrées dans la figure 5.6 pour chaque personne dans la base de données. Le nombre total des images de profondeurs utilisées est 2491 images. Nous étudions l'apport du descripteur LPQ multi-blocs en utilisant la transformation d'Euler pour générer les différentes poses. Nous examinons les résultats en utilisant différents nombres de blocs et différentes valeurs de rayon R du descripteur LPQ.

Nous présentons donc un système entraîné par des informations sur les personnes avec différentes poses sans l'utilisation de techniques d'estimation ou de correction de poses. Dans un premier temps, basé sur nos résultats obtenus dans le chapitre 4, nous pouvons fixer le nombre des blocs à 16*16. Ensuite, nous varions la valeur du rayon du descripteur LPQ. Cette dernière, est variée pour fixer la valeur optimale. Le tableau 5.2 présente les performances de vérification de notre système en fonction du rayon R de LPQ dans l'ensemble d'évaluation et de test.

Tableau 5.1 Protocole utilisé pour l'évaluation de variations de poses

Ensemble	Client	Imposteur
Apprentissage	700 images (1, 4, 8, 9, 10, β_1, β_9)	0 image
Evaluation	700 images (2, 6, 7, 14, 15, β_4, β_8)	221 images (1:15, β_5, β_7)
Test	700 images (3, 5, 11, 12, 13, β_2, β_{10})	170 images (1 :15, β_3, β_6)

Tableau 5.2 Performances de la vérification, effet du rayon de LPQ

Rayon de descripteur LPQ	Evaluation	Test	
	TEE (%)	HTER (%)	TV (%)
R=1	6.30	5.33	93.71
R=3	4.57	3.85	95.42
R=5	2.86	3.43	96.57
R=7	3.86	2.96	96.14
R=9	4.00	3.36	96.00
R=11	4.42	3.28	95.57
R=13	4.28	3.40	95.71
R=15	4.86	3.59	95.14
R=17	5.13	4.46	94.85

Chaque ligne du tableau précédent présente les performances sous forme de taux (TEE, HTER, et TV) pour une valeur de R donnée. Nous pouvons clairement constater que R=5 donne un taux de vérification égale à 96.57%. Les taux de vérification en fonction de la variation de rayon R changent autour de 96%. En présence des variations d'illuminations, des expressions faciales et avec une grande variation de poses, nous obtenons un bon résultat. La figure 5.7 présente une comparaison des taux de vérification en fonction de la variation du rayon R. Elle montre que les résultats obtenus sont globalement bons (taux de vérification entre 93.71% et 96.57%). Le changement de rayon R contribue significativement à l'amélioration des performances de notre système de vérification. En effet, plus le rayon R est grand (jusqu'à un certain seuil), plus le taux de vérification est élevé. Cela montre l'importance de la considération d'un grand voisinage de pixel lors de l'application du descripteur LPQ aux images de profondeur. En effet, le contraste au sein de l'information de profondeur dans un petit voisinage n'est pas suffisant pour une bonne discrimination à cause de l'aspect lisse de la surface faciale. Etant donné que les images faciales sont divisées en plusieurs blocs, nous menons des expérimentations sur l'effet du nombre de ces blocs, où la valeur de R est fixée à 5. Le tableau 5.3 présente les performances de vérification de notre système en fonction du nombre des blocs dans l'ensemble d'évaluation et de test. Il est bien évident que le meilleur nombre des blocs est 16*16 (256 blocs), 16 blocs selon x et 16 blocs selon y.

Tableau 5.3 Performances de la vérification, effet du nombre des blocs

Nombre de blocs (n*n)	Evaluation	Test	
	TEE (%)	HTEr (%)	TV (%)
n1=1	10.14	9.38	89.85
n2=4	5.14	4.27	94.85
n3=16	2.86	3.43	96.57
n4=32	4.47	5.51	94.57

La comparaison des performances du descripteur LPQ multi-blocs dans notre système de vérification par rapport au changement du nombre de blocs ou bien le nombre d'histogrammes est illustré dans la figure 5.7. Les résultats obtenus pour l'utilisation de l'image faciale sans subdivisions (le cas de nombre de blocs égal à 1) sont faibles avec un taux de vérification de 89.85%. Ceci s'explique par la raison que la représentation de visage globalement (sans subdiviser l'image) ne prend pas en considération l'information locale, ce qui nous conduit à un manque d'informations sur les caractéristiques locales (la bouche, le nez et les yeux) et donc une faiblesse significative dans le pouvoir de la discrimination. Le nombre des blocs utilisé influence significativement les performances de notre système. Les résultats obtenus sont globalement bons (le taux de vérification entre 94.57 % et 96.57 %). Plus le nombre des blocs par image est grand, plus le taux de vérification est élevé jusqu'à certain seuil (16*16). Après ce seuil idéal, on peut voir une redondance d'information qui affecte négativement les performances du système. Ces résultats sont confirmés avec les courbes ROCs. Comme le montre la figure 5.8 qui illustre la courbe ROC de notre approche lorsque le nombre des blocs 16*16 est toujours au-dessus des autres valeurs. Ce qui signifie que le taux de vérification est à chaque taux de fausses acceptations plus grand que le taux de vérification des autres nombre des blocs (1*1, 4*4, et 32*32).

L'étude expérimentale effectuée pour l'évaluation de notre système de vérification 3D de visage basée sur l'apprentissage par les angles d'Euler et à l'utilisation des images de profondeur avec les histogrammes du descripteur LPQ démontre les trois points suivants :

- Le choix du paramètre R est important pour le pouvoir de discrimination du descripteur LPQ.
- La subdivision de l'image d'entrée en plusieurs blocs est une opération très importante. La précision du descripteur LPQ par rapport au nombre des blocs utilisés lors de l'utilisation d'une valeur optimale de R (dans notre système R=5) est relativement stable.
- Un taux de vérification entre 98% et 100% en raison de la perte d'une partie de l'information à cause de la rotation de la tête avec un certain angle. Ceci peut être vu dans la figure 5.6.

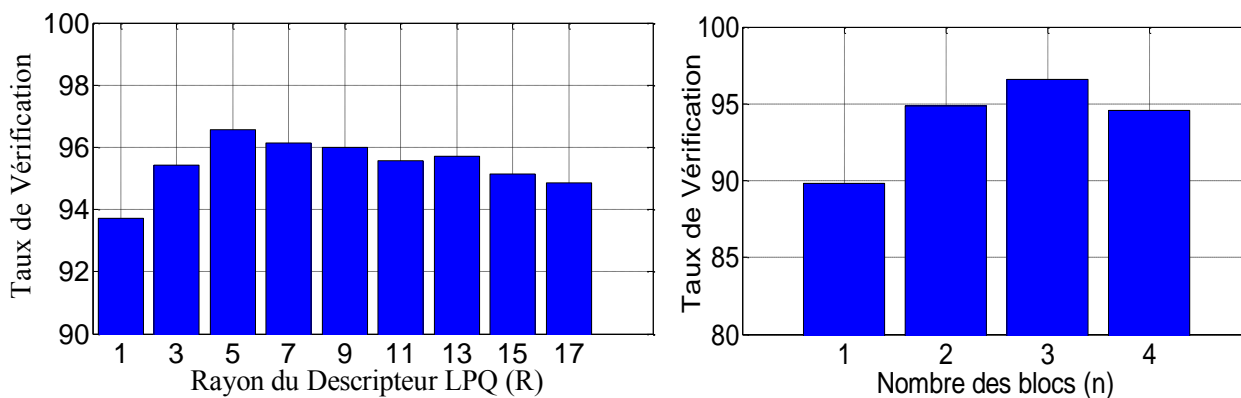


Figure 5.7 Variation de taux de vérification en fonction du rayon et du nombre des blocs LPQ

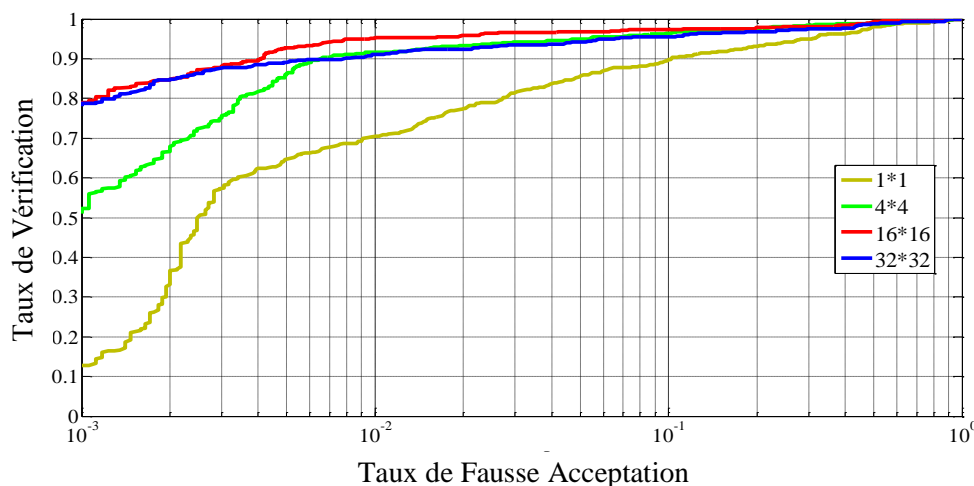


Figure 5.8 Les courbes ROCs pour différents nombres de blocs

5.4 Vérification 3D de visage basée sur les tenseurs

Généralement, le but de nombreux problèmes d'analyse de données (l'infographie, la vision par ordinateur, et l'apprentissage de la machine) est de trouver une représentation de données multi-variables appropriées qui facilitent l'analyse, la visualisation, la compression, l'approximation, et / ou l'interprétation des données. Cela se fait souvent par l'application d'une transformation appropriée à l'espace dans lequel les données d'observation résident. Plusieurs méthodes ont été développées pour calculer les transformations, résultant dans les représentations avec des hypothèses différentes et avec différentes propriétés d'optimalité définies en termes de réduction de la dimension (notre cas), l'intérêt et la pertinence statistique des composants qui en résultent, la simplicité de la transformation, et des différents critères d'applications liées. Les représentations qui sont dérivées par des transformations linéaires ont été préférées en raison de leur simplicité conceptuelle et computationnelle [157]. Par exemple, PCA et ICA sont des méthodes qui utilisent des transformations linéaires basées sur la représentation matricielle des données. Ces transformations sont limitées dans leur aptitude à faciliter l'analyse des données. Cependant, les transformations multilinéaires sont des

généralisations naturelles des transformations linéaires basées sur la représentation tensorielle. Ces dernières, fournissent beaucoup plus de puissance. Les transformations multilinéaires conduisent à des modèles génératifs qui capturent explicitement comment les données observées sont influencées par de multiples facteurs. En tant qu'applications dans l'approche tensorielle proposée, nous abordons des problèmes importants dans la reconnaissance faciale. En particulier, la représentation de l'image (histogramme des descripteurs locaux), les conditions d'éclairage, les expressions faciales et les variations de poses. Les deux catégories de représentation des images, tensorielle et matricielle sont illustrées dans les figures 5.9 et 5.10 respectivement.

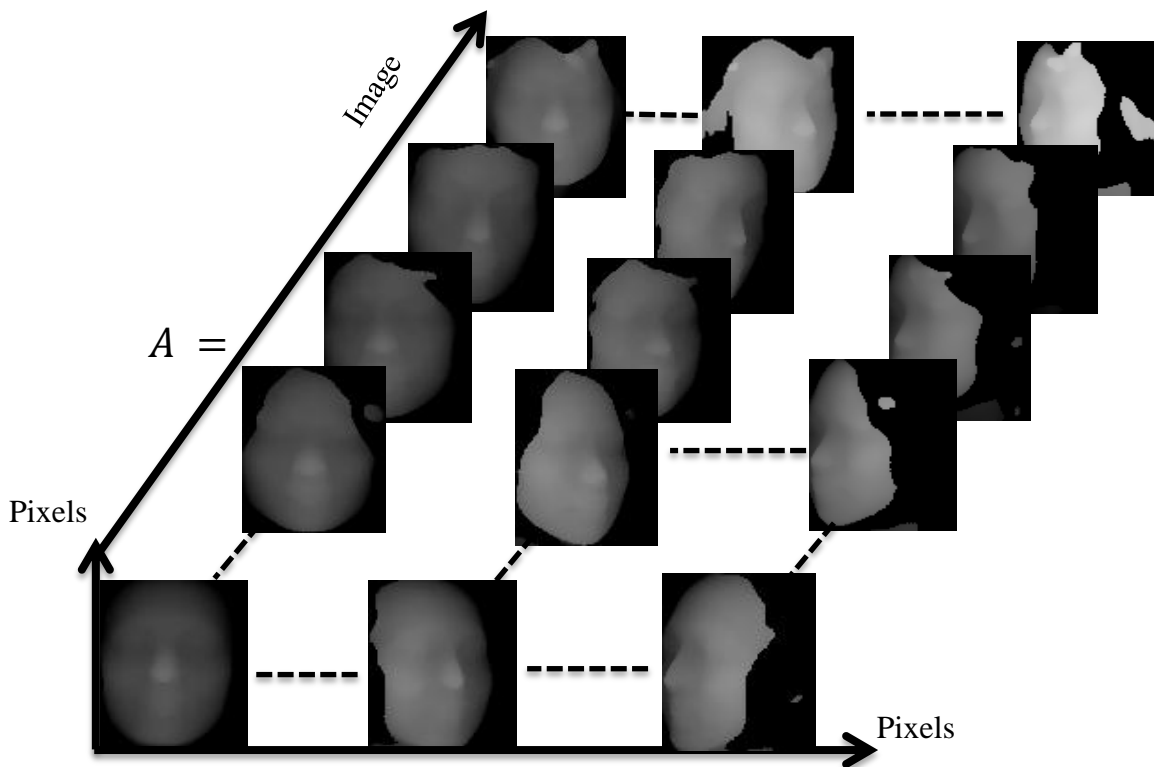


Figure 5.9 Représentation tensorielle d'ensemble d'images

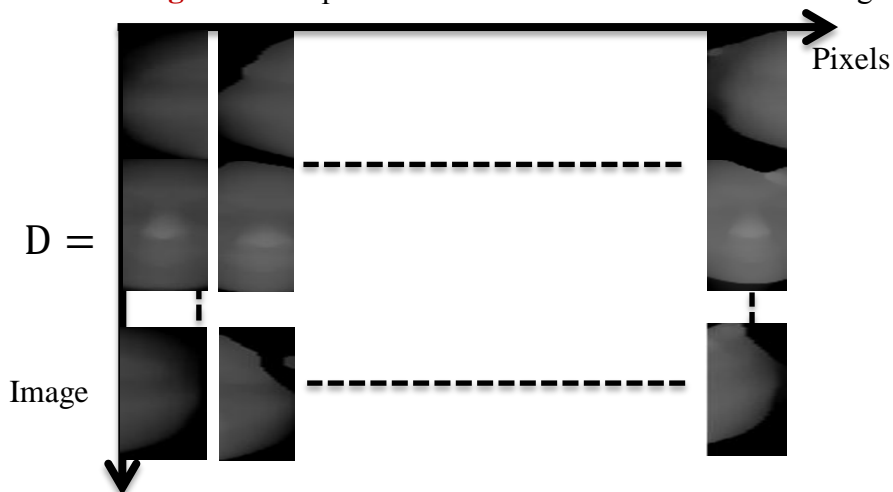


Figure 5.10 Représentation matérielle (vectorisation) d'ensemble d'images

Cette section vise à résoudre le problème de la reconnaissance de visage en présence d'une grande variété de poses. Notre système est basé sur l'analyse tensorielle et l'apprentissage multi-poses des images faciales à partir d'une seule image de profondeur en utilisant les angles d'Euler. Nous présentons les outils mathématiques d'algèbre multilinéaire qui seront utiles pour la modélisation des données par les tenseurs. En effet, nous nous intéressons à une génération multilinéaire de la méthode PCA+EFM en utilisant la décomposition en valeurs singulières pour la réduction des différents modes (trois modes) des tenseurs dans laquelle, le mode 1 est associé aux histogrammes LPQ de chaque personne, le mode 2 est associé aux échantillons (les personnes) et le mode 3 est associé aux différentes variations de poses ($\pm 60^\circ$).

5.4.1 Propriétés des tenseurs et modélisation multidimensionnelle

5.4.1.1 Définition d'un tenseur

Un tenseur est considéré comme une extension d'ordre supérieur d'un vecteur et d'une matrice. Un tenseur d'ordre N est noté de la façon suivante : $A \in \mathbb{R}^{I_1 \times \dots \times I_N}$ (5.6)

Le tenseur A est défini comme un tableau multidimensionnel avec N modes, les éléments du tenseur A sont notés par : $a_{i_1 \dots i_n \dots i_N}$ et I_N représente la dimension du n -mode ou $1 \leq i_n \leq I_N$, $1 \leq I_N \leq N$. L'espace du tenseur d'ordre N est composé de N mode de sous-espaces. Un ensemble de données comprenant N paramètres peut-être modéliser comme un tableau de N entrées, dans lequel, chaque mode du tenseur est associé à un paramètre différent. Dans le tableau 5.4, nous donnons des exemples de la modélisation tensorielle.

Tableau 5.4 Exemples de modélisations tensorielles

Exemple	Modélisations tensorielles
Longueur d'onde	Tenseur d'ordre 0
Coordonnées cartésiennes	Tenseur d'ordre 1
Image en niveau de gris	Tenseur d'ordre 2
Image RGB	Tenseur d'ordre 3
Séquence vidéo couleur	Tenseur d'ordre 4
Image hyper-spectrale	Tenseur d'ordre n (n : nombre de spectres)

5.4.1.2 Déploiement des tenseurs

La modélisation tensorielle nécessite des outils d'algèbre multilinéaire pour le traitement des données multidimensionnelles dans un mode privilégié [158]. À cette fin, les matrices de déploiement dans le n -mode sont choisies pour l'étude de tenseur selon un mode favorisé. Les « n -modes vecteurs » du tenseur A sont obtenus, en variant le $n^{\text{ième}}$ indice et en gardant les autres indices fixes. Le « n -mode matrices » $\mathbf{A}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \times \dots \times I_{n-1} \times I_{n+1} \times \dots \times I_N)}$ est formé en arrangeant tous les n -modes vecteurs séquentiellement ainsi que ses colonnes. La procédure pour former

les matrices n -modes à partir d'un tenseur d'ordre > 2 est appelée le déploiement du tenseur. La figure 5.11 illustre le déploiement d'un tenseur d'ordre 3 dans les différents modes.

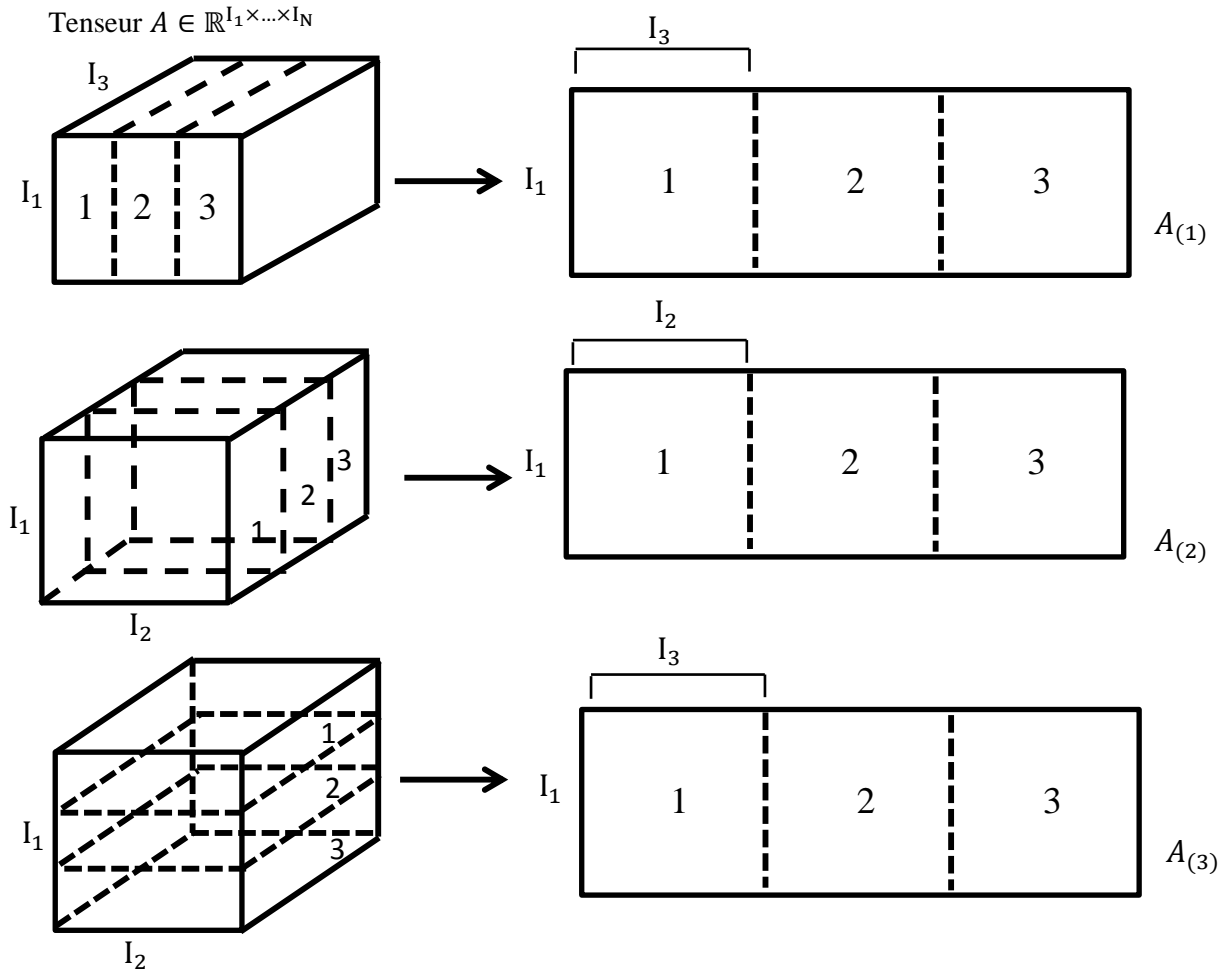


Figure 5.11 Le déploiement d'un tenseur d'ordre 3 dans les différents modes

5.4.1.3 Produit scalaire tensoriel

Considérons deux tenseurs A et $B \in \mathbb{R}^{I_1 \times \dots \times I_N}$. Le produit scalaire entre ces deux tenseurs est :

$$\langle A|B \rangle = \sum_{i_1, \dots, i_N} a_{i_1 \dots i_N} \cdot b_{i_1 \dots i_N} \quad (5.7)$$

Propriété :

$$\langle A|B \rangle = \text{tr} (\mathbf{A}_n \mathbf{B}_n^T) \quad (5.8)$$

Où $\text{tr}(\cdot)$ est l'opération trace d'une matrice.

Cas particulier :

La norme de Frobenius : $\|A\|$ est la norme de Frobenius d'un tenseur A , où :

$$\|A\|^2 = \sum_{i_1, \dots, i_N} a_{i_1 \dots i_N}^2 = \text{tr} (\mathbf{A}_n \mathbf{A}_n^T) \quad (5.9)$$

De cette relation, on peut définir la différence quadratique entre deux tenseurs A et $B \in \mathbb{R}^{I_1 \times \dots \times I_N}$ par $\|A - B\|^2$.

5.4.1.4 Produit extérieur

Le produit extérieur de plusieurs vecteurs $\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(N)}$ (de dimension I_1, \dots, I_N) appartenant aux espaces vectoriels $E^{(1)}, \dots, E^{(N)}$ respectivement, définit le tenseur $A \in \mathbb{R}^{I_1 \times \dots \times I_N}$ ou :

$$A = \mathbf{e}^{(1)} \circ \mathbf{e}^{(2)} \circ \dots \circ \mathbf{e}^{(N)} \quad (5.10)$$

L'élément $a_{i_1 \dots i_N}$ de tenseur A est défini par le produit :

$$a_{i_1 \dots i_N} = \prod_{n=1}^N e_{i_n}^{(N)} \quad (5.11)$$

Où $e_{i_n}^{(N)}$ est la $i_n^{\text{ième}}$ composante du vecteur $\mathbf{e}^{(N)}$

5.4.1.5 Opérateur n-mode produit

L'opérateur n-mode produit (\times_n) d'un tenseur $A \in \mathbb{R}^{I_1 \times \dots \times I_N}$ par une matrice $\mathbf{H} \in \mathbb{R}^{J_n \times I_n}$ est noté $A \times_n \mathbf{H}$ et donne un tenseur B tel que :

$$B = A \times_n \mathbf{H} \in \mathbb{R}^{I_1 \times \dots \times I_{n-1} \times J_n \times I_{n+1} \times \dots \times I_N} \quad (5.12)$$

Où chaque élément du tenseur est donné par :

$$B = (A \times_n \mathbf{H})_{i_1 i_2 \dots i_{n-1} j_n i_{n+1} \dots i_N} = \sum_{i_n=1}^{I_n} a_{i_1 i_2 \dots i_{n-1} i_n i_{n+1} \dots i_N} \cdot h_{j_n i_n} \quad (5.13)$$

Ce résultat, peut s'écrire avec la forme matricielle en utilisant les matrices dépliantes :

$$\mathbf{B}_n = \mathbf{H} \cdot \mathbf{A}_n \quad (5.14)$$

Où \mathbf{A}_n et \mathbf{B}_n représentent les matrices dépliantes des tenseurs A et B dans le n -mode. Ainsi le tenseur résultant des n -mode produits entre le tenseur $A \in \mathbb{R}^{I_1 \times \dots \times I_N}$ et les matrices $\mathbf{H}^{(n)} \in \mathbb{R}^{J_n \times I_n}$, $n = 1, \dots, N$ se définit par :

$$B = A \times_1 \mathbf{H}^{(1)} \dots \times_N \mathbf{H}^{(N)} \in \mathbb{R}^{J_1 \times \dots \times J_N} \quad (5.15)$$

Les éléments indice () du tenseur B sont donnés par :

$$b_{j_1 \dots j_N} = \sum_{i_1=1}^{I_1} \dots \sum_{i_N=1}^{I_N} a_{i_1 \dots i_N} h_{j_1 i_1}^{(1)} \dots h_{j_N i_N}^{(N)} \quad (5.16)$$

Où $i_k \in \{1, \dots, J_k\}$ et $J_k \in \mathbb{N}^*$, $k = 1, \dots, N$

5.4.2 Décomposition tensorielle par HOSVD

La Décomposition en Valeurs Singulières (Singular Value Decomposition, SVD) est un outil important de factorisation des matrices rectangulaires réelles ou complexes. La SVD est un algorithme qui permet d'exprimer une matrice $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2}$ sous forme d'un produit de trois matrices particulières \mathbf{U} , \mathbf{V} , et \mathbf{S} . La SVD de \mathbf{A} s'écrit sous la forme :

$$\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^T \quad (5.17)$$

dans laquelle :

$$\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_{I_1}] \quad (5.18)$$

\mathbf{U} est la matrice unitaire appartenant à $\mathbb{R}^{I_1 \times I_1}$ contenant les vecteurs singuliers gauches de la matrice \mathbf{A} . Les vecteurs $\mathbf{u}_k, k = 1, \dots, I_1$ sont les vecteurs propres de la matrice symétrique $\mathbf{A}\mathbf{A}^T$.

$$\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_{I_1}] \quad (5.19)$$

\mathbf{V} est la matrice unitaire appartenant à $\mathbb{R}^{I_2 \times I_2}$ contenant les vecteurs singuliers droits de la matrice \mathbf{A} . Les vecteurs $\mathbf{v}_k, k = 1, \dots, I_2$ sont les vecteurs propres de la matrice symétrique $\mathbf{A}^T\mathbf{A}$.

Enfin, la matrice pseudo diagonale \mathbf{S} appartenant à $\mathbb{R}^{I_1 \times I_2}$ est appelée la matrice « **noyau** », cette matrice contient les valeurs singulières de la matrice \mathbf{A} . Les matrices $\mathbf{A}\mathbf{A}^T$ et $\mathbf{A}^T\mathbf{A}$ ont le même rang (le rang d'une matrice \mathbf{A} égale au nombre de valeurs singulières \mathbf{S}_i non nulles) que la matrice \mathbf{A} , et ont les mêmes valeurs propres λ_i . Les valeurs singulières σ_k de la matrice \mathbf{A} sont

$$\sigma_i = \sqrt{\lambda_i}, i = 1, \dots, k \quad (5.20)$$

L'idée principale de la décomposition tensorielle consiste à trouver une bonne approximation d'un tenseur $A \in \mathbb{R}^{I_1 \times \dots \times I_N}$ par un autre tenseur \hat{A} de manière à minimiser un critère spécifique. Dans notre cas, le critère est la dimension des différents modes de tenseurs. Dans la représentation matricielle, la meilleure approximation au sens des moindres carrés d'une matrice \mathbf{A} est obtenue par la troncature de sa SVD [158]. C'est à dire qu'en conservant uniquement les k premiers vecteurs singuliers associés aux k plus grandes valeurs singulières de la matrice \mathbf{A} . La décomposition tensorielle par HOSVD est une généralisation de la SVD est proposée par *Lathauwer* [159] *et al* pour les tenseurs d'ordre supérieur. HOSVD est bien souvent utile pour pouvoir approximer un tenseur à l'aide de décomposition. Cet algorithme permet de déterminer les matrices orthogonales $\mathbf{U}^{(n)}$ où chaque matrice $\mathbf{U}^{(n)}, n = 1, \dots, N$ contient les K_n^1 vecteurs singuliers de la matrice dépliant \mathbf{A}_n , dans le mode n du tenseur A . L'algorithme HOSVD (voir tableau 5.5) est basé sur le modèle de décomposition tensorielle de Tucker3 [160]. La figure 5.12 illustre une décomposition de Tucker3 d'un tenseur d'ordre 3.

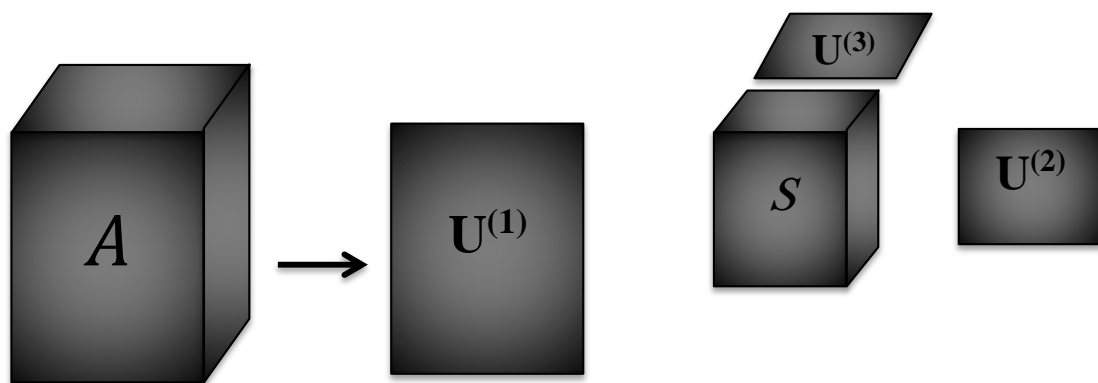


Figure 5.12 Décomposition de Tucker3 d'un tenseur d'ordre 3

Tableau 5.5 L'algorithme de décomposition en HOSVD**Algorithme : HOSVD**

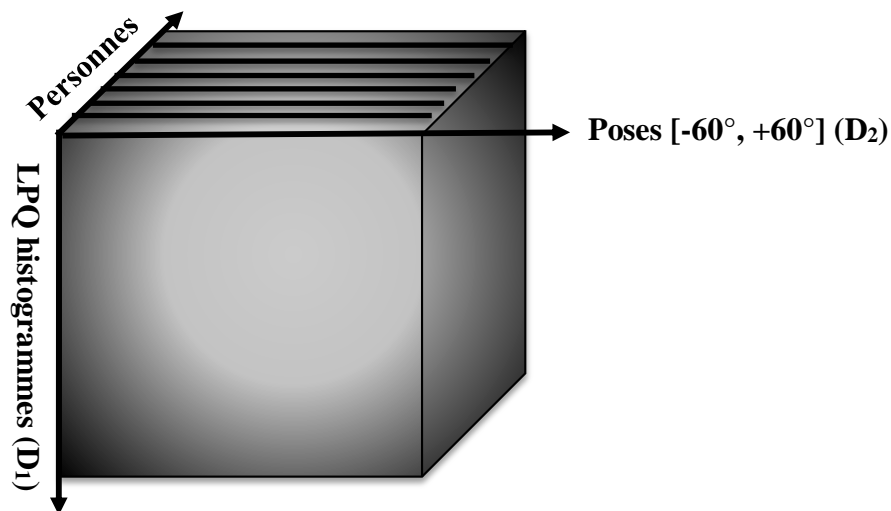
1. Entrées : Tenseur A , rangs n -modaux K_1, K_2, \dots, K_N
2. Pour $n = 1, \dots, N$, calcul des vecteurs singuliers n -modaux $\mathbf{U}^{(n)}$
 - (a) Déplier le tenseur A dans son mode n pour obtenir la matrice \mathbf{A}_n ;
 - (b) Effectuer la SVD de \mathbf{A}_n ;
 - (c) Extraire les matrices $\mathbf{U}^{(n)}$ formées par les K_n vecteurs singuliers associés aux plus grandes valeurs singulières de \mathbf{A}_n ;
 - (d) Calculer le tenseur noyau $S = A \times_1 \mathbf{U}^{(1)T} \times_2 \dots \times_N \mathbf{U}^{(N)T}$
3. Sorties : Tenseur noyau S et projecteurs n -modaux $\mathbf{U}^{(n)}, n = 1, \dots, N$.

5.4.3 Description de l'approche proposée

L'approche proposée est basée sur la création d'un tenseur d'ordre 3. La figure 5.13 illustre la forme du tenseur utilisé dans notre système de vérification de visage ou chaque mode de tenseur (D_1, D_2, D_3) représente un paramètre. Ces derniers sont définis comme suit :

- D_1 : Dimension du vecteur de caractéristiques LPQ histogrammes.
- D_2 : Poses, 13 angles de poses de la tête dans l'intervalle $[-60^\circ +60^\circ]$ avec un pas de 10° .
- D_3 : Les personnes, différentes personnes avec diverses variations d'expressions et d'illuminations.

La figure 5.13 illustre les différentes étapes de notre approche de vérification. Nous utilisons deux étapes de réduction de dimension, multilinéaire (HOSVD) et linéaire (PCA+EFM).

**Figure 5.13** Représentation du tenseur d'ordre 3 proposé

Après la détection et le prétraitement, les caractéristiques des histogrammes LPQ sont extraites et enregistrées sous forme d'un tenseur $A \in \mathbb{R}^{D_1 \times D_2 \times D_3}$ d'ordre 3. À partir de

l'HOSVD du A , en conservant dans le mode D_1 (les histogrammes LPQ) les K_n vecteurs singuliers associés aux plus grandes valeurs singulières de la matrice dépliant A . Nous obtenons un tenseur noyau réduit $S \in R^{L_1 \times L_2 \times D_3}$ $S = A \times_1 \mathbf{U}^{(1)T} \times_2 \mathbf{U}^{(2)T} \times_3 \mathbf{U}^{(3)T}$ où $(L_1 \times L_2 \ll D_1 \times D_2)$. La matrice dépliant S du tenseur S est de taille $L \times D_3$. Les vecteurs colonnes de la matrice S représentent les personnes et ses vecteurs lignes représentent les caractéristiques des personnes sous différentes variations d'illuminations, d'expressions et de poses. Cet espace de caractéristiques est ensuite réduit par la matrice de projection PCA+EFM (voir la section 3.5 du chapitre 3) pour augmenter l'efficacité de la méthode grâce à l'aspect classe de la méthode EFM. Ensuite, une matrice d'apprentissage réduite M_G de taille $k \times D_3$ ($k \ll L$) est obtenue où k est le nombre de caractéristiques finales utilisées pour la comparaison. Cette matrice, contient des vecteurs de caractéristiques significatives pour chaque sujet.

Dans la phase de test, l'image d'entrée est traitée de la même façon que les images d'apprentissage et projetée tout d'abord dans le sous-espace HOSVD et concaténée sous forme d'un vecteur de taille $L_1 \times L_2$. Ce vecteur, est ensuite projeté dans le sous-espace PCA+EFM. Enfin, nous obtenons un vecteur de caractéristiques de taille réduite k pour le sujet (candidat) de test. Ce dernier est comparé avec la matrice M_G afin d'arriver à une décision d'accepter ou rejeter le candidat. La figure 5.14 illustre une présentation générale du notre système de vérification proposée.

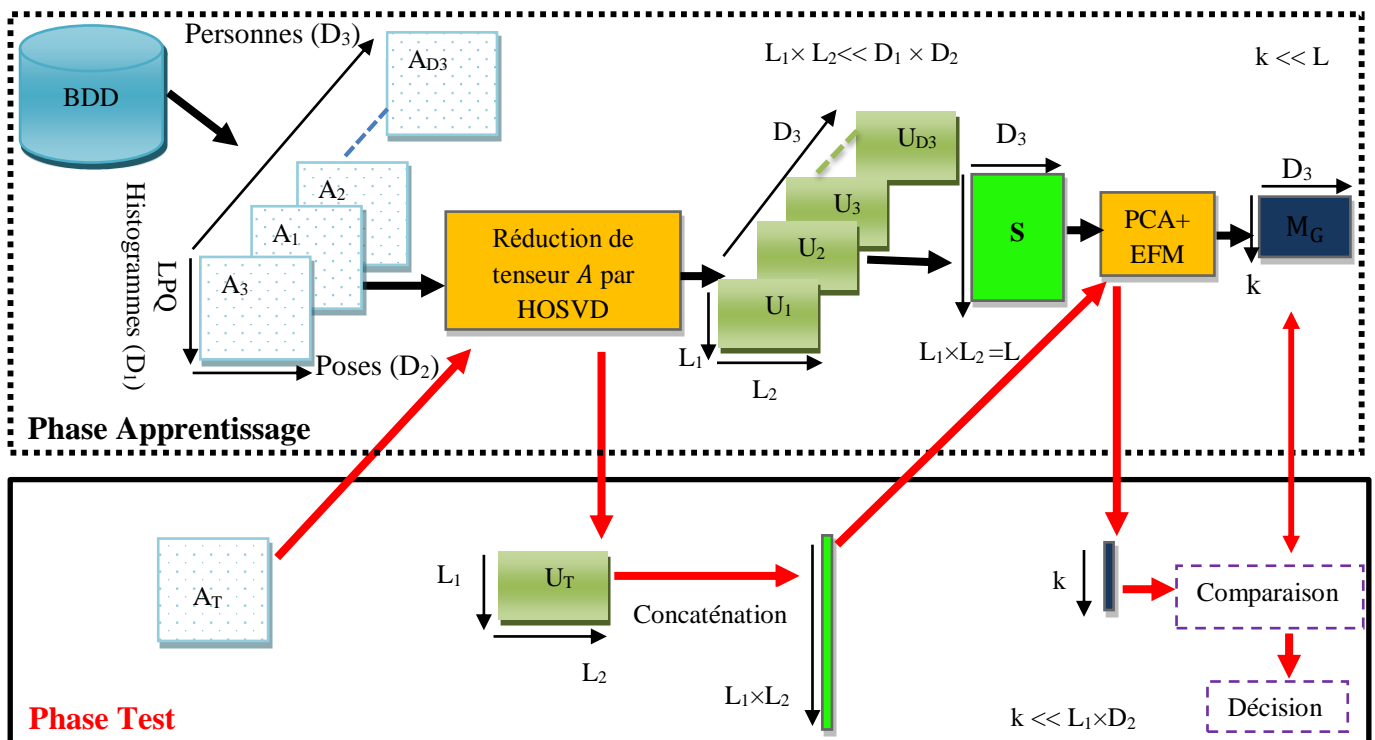


Figure 5.14 Présentation générale de l'approche proposée

5.4.4 Résultats et discussion

Une série d'expérimentations sont présentés afin d'évaluer l'algorithme proposé pour la vérification du visage 3D basé sur la représentation des données avec les tenseurs d'ordre supérieur (troisième). Tout d'abord, nous décrivons brièvement les ensembles de données dans la base de données CASIA3DV4 et le protocole sur lequel les expériences ont été réalisées. Ensuite, nous évaluons la vérification du visage 3D pour trois méthodes : (1) Concaténation des poses PCA+EFM, (2) HOSVD et (3) HOSVD+PCA+PFM. Les résultats sont comparés pour prouver l'efficacité de l'utilisation des deux étapes de projection, la première multilinéaire (HOSVD) et la deuxième linéaire (PCA+EFM).

La base de données CASIA3DV4 est utilisée pour tester les performances. À cette fin, nous avons utilisé le protocole présenté dans le tableau 5.6.

Tableau 5.6 Protocole utilisé pour l'évaluation des trois méthodes

Dataset	Client	Imposteur
Apprentissage	6500 images (65 images/personne)	0 image
Evaluation	6500 images (65 images/personne)	2535 images (195 images/personne)
Test	6500 images (65 images/personne)	1950 images (195 images/personne)

Chaque personne est représentée par 195 images de profondeur. Nous avons utilisé 23985 images de 123 personnes avec différentes variations mentionnées ci-dessus. Ce protocole est basé sur le protocole présenté dans le tableau 4.1 pour les poses frontales avec un angle de 0° . Cependant, pour chaque image de personne avec la pose frontale (de 1 à 15), 12 images de profondeur avec des variations de poses sont générés en utilisant la transformation d'Euler. Chaque visage subit une rotation dans l'intervalle des angles $[+60^\circ, -60^\circ]$ avec un pas de 10° . Basé sur nos résultats présentés dans le chapitre 4 et la section 5.3, les histogrammes de descripteur LPQ multi-blocs sont utilisés comme un vecteur de caractéristiques. Pour la méthode linéaire de réduction de dimension PCA+EFM, les 12 images de poses pour chaque personne sont concaténées avec l'image de la pose frontale, et une matrice d'apprentissage géante est générée. Pour la méthode multilinéaire HOSVD, les données sont représentées sous forme d'un tenseur d'ordre 3. Où, la dimension 1 représente les histogrammes LPQ, la dimension 2 représente les poses et la dimension 3 représente les personnes avec différentes variations d'illuminations et d'expressions faciales.

Pour chaque expérience, la dimension des vecteurs de caractéristiques est variée jusqu'à l'obtention des meilleures performances. Dans la stratégie proposée " HOSVD+PCA+EFM", PCA est utilisée pour réduire la taille des caractéristiques des histogrammes après l'extraction

de caractéristiques multilinéaires par HOSVD, de sorte que la matrice de dispersion S_w (within matrix) est inversible (EFM est plus stable dans ce cas).

Les performances de vérification durant l'évaluation et le test ainsi que les dimensions correspondantes pour les trois méthodes sont présentées dans les tableaux (5.7, 5.8 et 5.9).

La dimension D_1 du tenseur varie entre 0 et 600, les meilleurs résultats se trouvent avec les dimensions de 400 à 600 pour les méthodes PCA+EFM. Tandis que pour HOSVD+PCA+EFM les meilleures performances sont obtenues avec un nombre de caractéristiques (Nbr caract) inférieur à 100. Comme cela est visible d'après le tableau 5.9 et la figure 5.16 (b), HOSVD+PCA + EFM donne les meilleurs taux de vérification ainsi qu'une bonne stabilité durant la phase de test avec un taux d'égale erreur faible égal à 1.40% et un taux de vérification élevé de 98.60%. Ces excellents résultats, démontre la robustesse de notre méthode proposée aux différents facteurs de variabilité (illuminations, expressions, poses $\pm 60^\circ$).

Tableau 5.7 Performances de la vérification PCA+EFM

Nbr caract	Evaluation	Test	
	TEE (%)	HTER (%)	TV (%)
10	2.99	6.19	92.40
20	1.79	4.50	94.40
30	1.24	4.04	95.26
40	1.41	4.26	94.93
50	1.38	3.89	95.40
60	1.24	3.67	95.62
70	1.23	3.46	95.93
80	1.24	3.51	95.88
90	1.21	3.52	95.88
100	1.20	3.53	95.86

Tableau 5.8 Performances de la vérification HOSVD

D1 Tenseur	Evaluation	Test	
	TEE (%)	HTER (%)	TV (%)
10	12.39	10.28	87.60
20	10.80	9.40	89.20
30	9.56	8.72	90.40
40	8.95	8.06	91.00
50	8.58	7.61	91.40
60	8.20	7.50	91.80
70	7.83	7.14	92.20
80	7.78	7.00	92.20
90	7.60	6.90	92.40
100	7.39	6.84	92.60
200	7.01	6.35	93.00
400	6.80	6.05	93.20
600	6.78	5.93	93.20

Tableau 5.9 Performances de la vérification avec HOSVD+PCA+EFM

Nbr caract	Evaluation	Test	
	TEE (%)	HTER (%)	TV (%)
10	6.63	6.25	93.40
20	4.19	3.56	95.80
30	2.60	2.36	97.40
40	2.40	1.87	97.60
50	2.20	2.08	97.80
60	1.99	1.74	98.00
70	2.00	1.71	98.00
80	2.43	1.51	98.00
90	1.40	1.03	98.60
100	1.79	1.28	98.20

Les taux de vérification et les taux d'égalité erreur par rapport aux variations de la dimension du tenseur pour la méthode HOSVD sont indiqués sur la figure 5.15, dans laquelle la dimension (l'axe horizontal) désigne le nombre de vecteurs lignes dans les matrices de projection après le déploiement du tenseur.

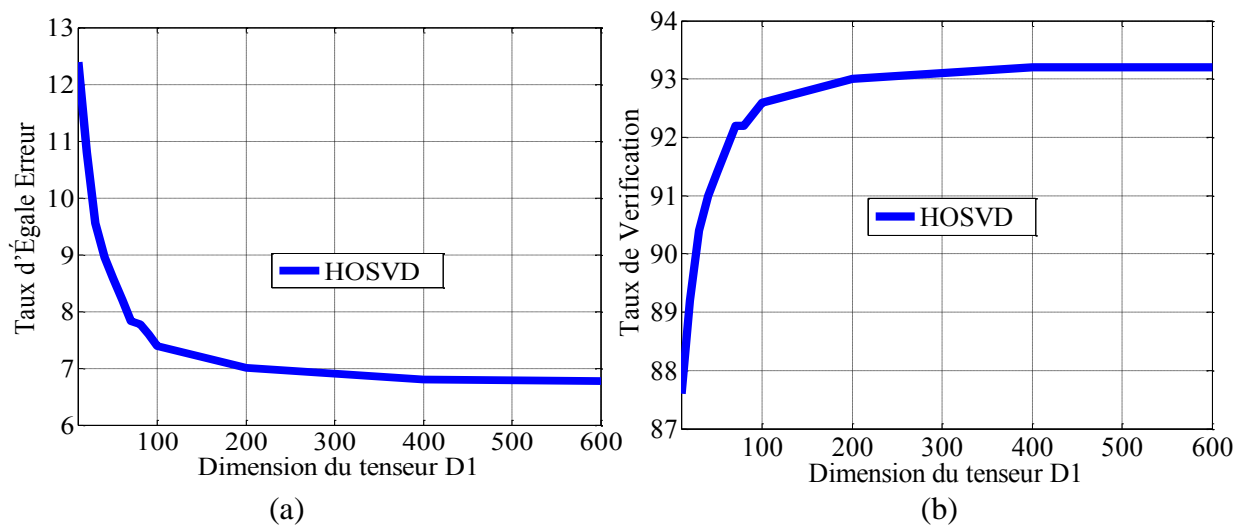


Figure 5.15 Performances de vérification avec HOSVD ;
 (a) TEE en fonction de la dimension de tenseur D1,
 (b) TV en fonction de la dimension de tenseur D1

Les taux de vérification et les taux d'égalité erreur par rapport aux variations du nombre de caractéristiques EFM pour la méthode PCA+EFM et HOSVD+PCA+EFM sont indiqués sur la figure 5.16, dans laquelle la dimension (l'axe horizontal) désigne le nombre de vecteurs lignes dans les matrices de projection.

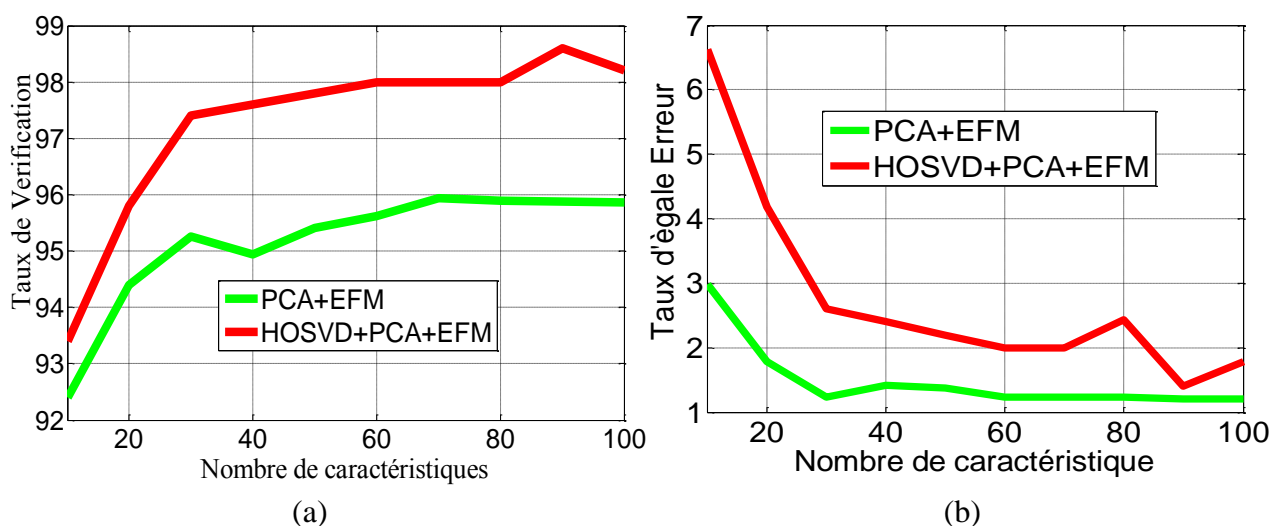


Figure 5.16 Performances de vérification avec PCA+EFM et HOSVD+PCA+EFM ;
 (a) TEE en fonction du nombre de caractéristiques,
 (b) TV en fonction du nombre de caractéristiques

D'après les résultats expérimentaux, il est évident que les méthodes multilinéaires basées sur les tenseurs sont plus stables entre les ensembles d'évaluation (TEE) et de test (TV). La méthode linéaire PCA+EFM donne de bons résultats durant l'ensemble d'évaluation (TEE entre 2.99 % et 1.20 %). Malgré cela, le TV n'est pas amélioré dans l'ensemble de test (TV autour de 95 %). Par exemple un TEE de 1.20% est obtenu dans l'ensemble d'évaluation alors que dans l'ensemble de test un TV de 95.86 % est atteint. Contrairement, pour la méthode multilinéaire lorsque nous utilisons la méthode HOSVD, les résultats entre l'ensemble d'évaluation et de test sont plus compatibles. Prenons comme exemple la méthode combinée HOSVD+PCA+EFM, un TEE de 1.40 % est obtenu dans l'ensemble d'évaluation et dans l'ensemble de test un TV élevé de 98.60 % est atteint. La stabilité est illustrée sur la figure 5.16. Le tableau 5.10 résume l'ensemble des meilleures performances relatives à chaque méthode ainsi que le temps de calcul pour le test d'une personne pour chaque méthode.

Tableau 5.10 Comparaison des performances avec différentes méthodes

Catégorie	Méthodes utilisées	Evaluation	Test		
		TEE (%)	HTER (%)	TV (%)	TC(s)
Multilinéaire	HOSVD	6.80	6.05	93.20	0.61
Linéaire	PCA+EFM	1.23	3.46	95.96	0.56
Multilinéaire +Linéaire	HOSVD+PCA+EFM	1.40	1.03	98.60	0.63

Nous soulignons que le temps de calcul est considérablement réduit ce qui rend notre méthode facile à mettre en œuvre en pratique. Lorsque nous comparons les résultats en terme de temps de calcul, la PCA+EFM donne un TV=95.96 % avec un temps de calcul égal à 0.56 (s).

La méthode HOSVD+PCA+EFM donne un TV= 98.60 % avec un temps de calcul égal à 0.61 (s). Alors que, HOSVD+PCA+EFM atteint un gain de TV de 2.4 % avec une perte de 0.05 (s) en termes de temps de calcul. Cela est un autre avantage pour notre méthode. Grâce à nos expérimentations dans la deuxième partie de ce chapitre, nous pouvons dire que l'idée de combiner deux étapes de réduction de données multilinéaires en utilisant la méthode HOSVD suivie par une méthode linéaire basée sur PCA+EFM donne de meilleures performances avec une bonne stabilité par rapport à l'utilisation de chaque méthode de façon indépendante.

La comparaison des trois méthodes en termes de courbes ROCs est présentée dans la figure 5.17, qui illustre la probabilité d'une vérification correcte par rapport au taux de fausse acceptation.

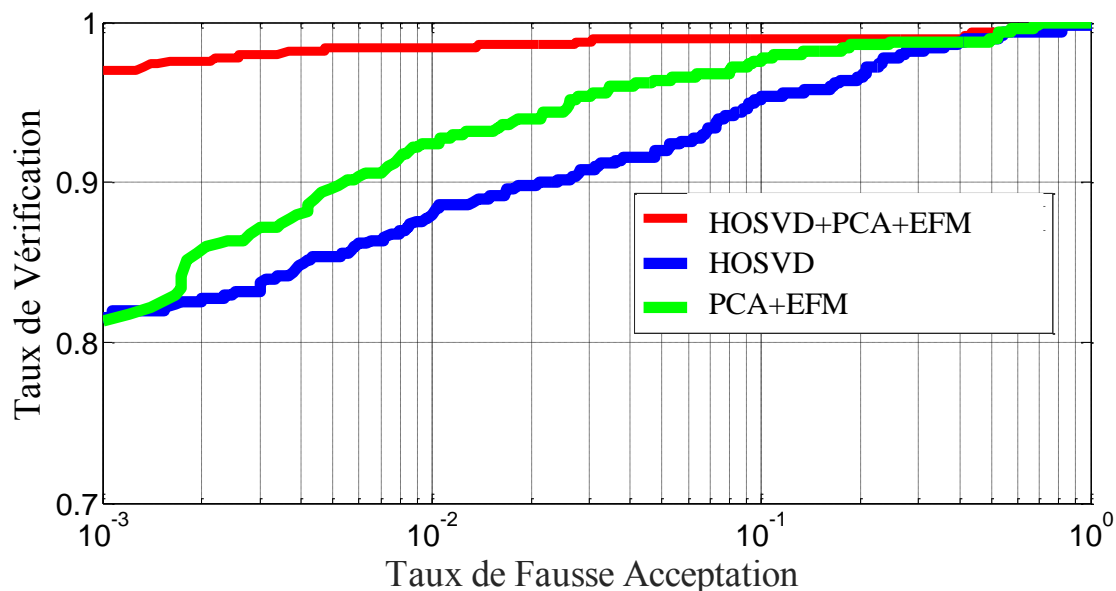


Figure 5.17 Courbes ROCs des trois méthodes

La courbe ROC de notre approche HOSVD+PCA+EFM a des performances meilleures que les autres méthodes utilisées et c'est la plus stable, ce qui signifie que le taux de vérification de notre algorithme est à chaque taux de fausse acceptation plus grand que le taux de vérification des autres méthodes.

5.5 Conclusion

Dans ce chapitre, nous avons présenté un nouveau cadre pour la vérification de visage basé sur l'apprentissage par les angles d'Euler en utilisant les images de profondeur avec les histogrammes de descripteur LPQ multi-blocs. Une matrice de rotation en 3D a été utilisée pour générer plusieurs poses qui sont utilisées pour l'apprentissage et le test. Grâce à l'utilisation de descripteurs LPQ, l'information de profondeur devient plus discriminante. Deux paramètres importants sont évalués, le rayon R et le nombre des histogrammes qui sont représentés par le

nombre de blocs utilisés. Les résultats obtenus sur la base CASIA3DV4 ont démontré tout l'intérêt de cette approche pour résoudre le problème de vérification de visage 3D en présence d'une grande variation de poses avec une méthode simple sans avoir recours à l'utilisation de correction ou d'estimation de poses. Notre approche, présente un autre avantage : nous pouvons utiliser les bases de données de visages 3D avec des poses frontales et nous générons les différentes poses pour chaque personne pour évaluer le système en présence de poses. Dans la deuxième partie du chapitre, nous avons mis au point une nouvelle approche de vérification de visage basée sur l'analyse tensorielle. Les défis : illuminations, expressions et poses sont considérés comme des problèmes d'algèbre multilinéaire dans laquelle l'ensemble des images de données est représenté comme un tenseur d'ordre 3. Ce dernier est décomposé en utilisant la méthode proposée HOSVD+PCA+EFM afin de réduire l'espace de caractéristiques. Un taux de vérification élevé de 98,60% a été obtenu, lorsque nous avons utilisé 23985 images de profondeur de 123 personnes différentes. Le chapitre suivant est consacré à la conclusion de ce mémoire de thèse en résumant les points principaux de nos contributions et quelques perspectives que nous envisageons d'explorer suite à notre travail.

Chapitre 6

Conclusion et Perspectives

6.1 Conclusion générale	110
6.2 Perspectives	110

6.1 Conclusion générale

Dans cette thèse, nous nous sommes intéressés aux problèmes de la reconnaissance automatique de visages en présence de différentes variations. Nous avons décrit d'une façon générale le domaine de la biométrie et les différentes modalités utilisées pour reconnaître les personnes en utilisant la machine. Nous avons présenté le procédé du système de reconnaissance faciale et les différents problèmes et défis scientifiques rencontrés dans le monde réel. Nous avons présenté les bases de données de visage 2D et 3D les plus utilisées dans la littérature pour l'évaluation des différents algorithmes de reconnaissance de visage ainsi que les critères d'évaluation et de mesure des performances pour les deux modes de reconnaissance vérification et identification.

Un état de l'art des méthodes récentes et travaux antérieurs sur la détection et la reconnaissance de visage 3D est mis au point. Ceci nous a permis de choisir la méthodologie adéquate à la problématique et faire un choix judicieux pour la conception du système d'analyse d'images de visage en présence d'expressions et de poses.

La détection est une étape primordiale dans les systèmes qui nécessitent une analyse de l'image faciale telle que le système de vérification du visage. Plus spécifiquement, nous avons abordé les méthodes de détection 3D basées sur le bout du nez comme un point de référence.

Dans un premier temps, nous avons proposé un système automatique de reconnaissance de visage en utilisant les images 3D de profondeur et 2D d'intensité couleur. Dans notre système, la région faciale a été détectée en se basant sur une nouvelle méthode proposée appelée la segmentation de la zone du visage avec les courbes de projection intégrale IPC suivie par la localisation de bout du nez dans laquelle certaines informations supplémentaires inutilisables sont éliminées et la zone du visage est localisée avec une grande précision. Notre méthode de détection ne nécessite aucune intervention manuelle, elle est entièrement automatique. Nous avons comparé deux classificateurs : la mesure de distance avec la corrélation normalisée et le SVM. Par la suite, nous avons adopté l'utilisation de SVM dans le reste de nos expérimentations, en raison de leur efficacité et leur précision.

La comparaison de nos résultats avec les méthodes existantes a démontré que nous obtenons des résultats prometteurs. Notre procédé donne une grande précision et une robustesse aux variations d'illuminations, d'expressions faciales et aux petites variations de poses inférieure à 35 °.

Pour parvenir à notre objectif, nous avons analysé les avantages et les inconvénients des méthodes actuellement utilisées ce qui nous a permis d'explorer plusieurs pistes et concevoir

de nouvelles stratégies de reconnaissance. Nous avons proposé plusieurs méthodes ciblées à des représentations robustes du visage contre la diversité des facteurs difficiles qui sont mentionnés ci-dessus. Pour cela, cinq parmi les descripteurs locaux les plus efficaces LBP, LPQ, BSIF, TPLBP et FPLBP ont été utilisés afin de fournir plus de puissance discriminative pour les images de profondeur. L'image d'entrée est subdivisée en plusieurs petits blocs rectangulaires et les histogrammes de chaque bloc ont été extraits. Ensuite, tous les histogrammes des différents blocs sont concaténés dans un seul vecteur de caractéristiques. Ce dernier, est considéré comme une signature biométrique efficace pour la représentation des personnes enregistrées dans la base de données. De cette manière, l'information locale est représentée (yeux, nez, bouche, front...) à partir de l'image faciale d'une façon discriminante. Des expériences sur les combinaisons de ces descripteurs au niveau des caractéristiques à l'aide d'une concaténation simple des histogrammes sont effectuées. Afin de réduire la grande dimensionnalité des vecteurs caractéristiques, trois méthodes de réduction de la dimensionnalité sont comparées la PCA, OLPP et PCA + EFM.

Les descripteurs BSIF et LPQ conduisent à de hautes performances de vérification ; nous obtenons un taux de vérification de 97,26% et 98,18% pour BSIF et LPQ respectivement. Les histogrammes LPQ explorent plus de caractéristiques discriminantes pour la vérification et surpassent toutes les autres caractéristiques des différents descripteurs (TV= 98.18 % et un HTER=0.90%). Nous attribuons ce résultat au fait que les illuminations et les variations d'expressions faciales n'ont aucun effet sur le domaine fréquentiel qui est représenté par le LPQ. Basés sur ces résultats, nous avons combiné le descripteur LPQ avec tous les autres descripteurs. En plus, nous avons également démontré à travers une étude comparative que la méthode combinée LPQ_FPLBP basée sur PCA+EFM est le meilleur chemin pour résoudre le problème des variations d'illuminations et d'expressions faciales grâce à ces performances élevées avec un TV égale à 98.22% et un HTER égale à 0.88%.

Nous avons souligné aussi le rôle important du processus de la réduction de dimensions. Nos résultats expérimentaux montrent que la méthode PCA + EFM a obtenu un gain plus de 10% dans le TV et un gain de 5,8 % dans le TEE par rapport à la méthode de base PCA. D'autre part, nous pouvons voir qu'il n'y a pas un gain de performance significatif dans les méthodes combinées. Le descripteur LPQ dépasse certaines méthodes combinées telles que : LPQ_TPLBP, LPQ_BSIF et LPQ_LBP. Cela est dû au fait que ces descripteurs partagent beaucoup de caractéristiques similaires qui ont conduit à un affaiblissement du processus de classification.

Les variations de poses sont des difficultés insurmontables aux SRV. Pour cela, nous avons consacré une majeure partie de nos travaux à résoudre ces problèmes.

Un nouveau cadre pour la vérification de visage basée sur l'apprentissage par les angles d'Euler est conçu en utilisant les images de profondeur avec les histogrammes de descripteur LPQ multi-blocs. Une matrice de rotation en 3D orthogonale de déterminant 1 a été utilisée pour générer plusieurs angles de rotation de la tête qui sont utilisées pour l'apprentissage et le test. L'avantage de l'approche proposée est que nous pouvons évaluer un système de vérification de visage 3D en présence de poses même dans les bases de données qui ne contiennent pas des variations de poses.

Grâce à l'utilisation de descripteur LPQ multi-blocs, l'information de profondeur devient plus discriminante, dans lequel deux paramètres très importants sont évalués : le rayon R et le nombre des histogrammes qui sont représentés par le nombre de blocs utilisés.

Les méthodes basées sur l'apparence, ont attiré beaucoup d'attention en vision par ordinateur. L'analyse traditionnelle basée sur l'apparence employait l'analyse en composantes principales pour modéliser la variation apparente dans des ensembles d'image d'apprentissage typiques en utilisant la décomposition en valeurs singulières de la matrice connue (SVD) et la réduction de dimension par omission d'ordre supérieur de vecteurs singuliers.

La variation multifactorielle de l'ensemble des images faciales provoque de graves difficultés dans la reconnaissance conventionnelle du visage à base de méthodes d'apparence. Nous avons introduit un puissant cadre : le tenseur pour l'analyse et la reconnaissance basée sur l'apparence des images qui élégamment et efficacement traite de la variation multifactorielle inhérente à l'apprentissage de l'image.

Notre approche exploite l'algèbre multilinéaire (l'algèbre des tenseurs d'ordre supérieur). L'HOSVD est une extension multilinéaire naturelle de la SVD aux tenseurs, connue aussi sous le nom SVD N-mode, pour analyser des ensembles de visage dans lesquels les images multiples modes sont autorisées à varier.

La méthode sépare et représente chacun des différents modes sous-jacents à l'apprentissage d'images faciales, permettant la réduction de la dimension spécifique de mode.

L'image du visage est représentée par un ensemble de vecteurs de coefficients, chaque mode constituant le facteur personne, point de vue, l'éclairage, l'expression, etc. L'approche tenseur algébrique est prometteuse pour la reconnaissance faciale dans les milieux non contrôlés.

Dans ce mémoire nous avons présenté un système de reconnaissance faciale multidimensionnelle, nous avons implémenté une technique pour la réduction des données pour l'analyse biométrique multidimensionnelle du visage.

Une partie de ce mémoire porte sur l'étude des tenseurs pour la conception d'une approche d'apprentissage de sous-espaces multilinéaires pour la reconnaissance faciale humaine basée sur l'apparence.

L'avantage majeur dans l'analyse multidimensionnelle est le fait que les données sont prises directement sous la forme de tenseurs d'ordre supérieure comme entrée. Ces méthodes permettent aux relations entre les dimensions d'une représentation tensorielle d'être caractériser efficacement. De nombreuses expériences montrent que nos méthodes ne sont pas seulement plus efficaces mais aussi plus performantes. Les analyses théoriques d'algèbre multilinéaire ont été présentées pour explorer les propriétés des tenseurs.

Notre application utilisant les tenseurs a été organisée comme suit :

La première phase est réalisée à base des histogrammes LPQ multi-blocs pour la construction des caractéristiques.

La deuxième phase utilise trois méthodes de réduction **PCA+EFM**, **HOSVD**, **HOSVD+PCA+EFM**. Notons que notre travail est validé par les taux de performance HTER, TEE, et TV.

Comme nous l'avons mentionné plus haut une méthode appelée HOSVD + PCA + EFM a été proposée pour la réduction de la dimensionnalité des tenseurs multidimensionnels.

La PCA + EFM est ajoutée après l'obtention des caractéristiques par la projection des tenseurs multilinéaires afin d'améliorer les performances. La comparaison des trois méthodes : PCA + EFM, HOSVD et HOSVD + PCA + EFM est donnée pour prouver l'efficacité de l'utilisation de deux étapes de réduction de la dimension, la première est multilinéaire basée sur HOSVD et la deuxième est une méthode linéaire utilisant PCA + EFM.

À ce moment-là, une de nos nouvelles contributions était la généralisation multilinéaire de la méthode PCA+EFM. Nous avons obtenu un taux de vérification élevé égal à 98.60 % avec la prise en compte des variations d'illuminations, d'expressions et une grande variation de poses ($\pm 60^\circ$ avec des pas de 10°) dans la base de données CASIA3DV4. Nous avons utilisé **23985** images de profondeur de 123 personnes différentes.

Le temps de calcul (0.63 s) est considérablement réduit ce qui rend notre méthode facile à mettre en œuvre en pratique. Notre méthode HOSVD+PCA+EFM atteint un gain de TV de **2.4 %** avec une perte de **0.05 (s)** par rapport à la méthode matricielle PCA+EFM. La réduction multilinéaire (HOSVD) suivie par une méthode linéaire (PCA+EFM) donne des meilleures performances avec une bonne stabilité.

6.2 Perspectives

D'après les résultats livrés par les propositions dans cette thèse, il y a beaucoup de questions sans réponses. Le système de reconnaissance faciale avec ces différentes étapes est une histoire sans fin où la perfection est difficile à trouver, en particulier dans des conditions non contrôlées. À notre modeste avis, l'extension possible de notre travail s'articule autour de quatre axes :

A) La sélection des caractéristiques basée sur les algorithmes d'optimisation

Nous nous concentrerons sur l'exploitation de la région nasale pour la correspondance (car cette région est la plus stable à la variation d'expressions) par les méthodes de sélection de caractéristiques discriminantes basées sur les algorithmes d'optimisation : optimisation par essaim de particules (Particle Swarm Optimization, PSO), algorithme génétique (Genetic Algorithms, GA) et les nouvelles méthodes méta heuristiques.

B) La fusion des descripteurs locaux avec une méthode statistique

La fusion des descripteurs locaux parfois conduit à un affaiblissement du processus de classification, parce qu'ils partagent beaucoup de caractéristiques similaires, alors dans ce cas nous avons une redondance d'informations faciales locales. Ce que nous avons observé lors de la fusion de LPQ avec BSIF. Nous arrivons à un bon résultat lors de l'utilisation du descripteur LPQ séparément par rapport à l'utilisation de la méthode de fusion LPQ_BSIF. Pour cela, nous trouvons intéressante l'idée de se concentrer sur le développement d'une méthode statistique de fusion des descripteurs en tenant compte de la relation de dépendance entre les descripteurs avec laquelle ils sont fusionnés. Une bonne compréhension de ce point devrait permettre de mieux combiner les caractéristiques des deux descripteurs de façon à bénéficier pleinement de leur fusion.

C) L'application de nos méthodes pour la classification de texture

Une question reste posée : est-il possible d'appliquer la méthode proposée (les descripteurs locaux multi-blocs basés sur HOSVD+PCA+EFM) dans ce genre de tâche ? Ça serait important si la réponse est affirmative.

D) La détection de visage basé sur des descripteurs plus efficaces LPQ et SVM

L'amélioration de la détection de visage est un autre chemin de recherche très motivant, comme indiqué par nos résultats : le descripteur LPQ multi-blocs est une bonne méthode pour représenter les caractéristiques de visage invariantes au changement d'expressions faciales et

d'illuminations. Inspiré par le détecteur Viola-Jones qui est basé sur une représentation de l'image sous forme d'intégrale pour extraire les caractéristiques d'Haar afin d'augmenter la discrimination entre le visage et le non-visage. Ensuite, l'apprentissage de ce détecteur est basé sur l'algorithme Adaboost afin de former des modèles de classification. Nous envisageons de développer un détecteur de visage basé sur les caractéristiques invariantes de descripteur LPQ suivi par le classificateur SVM à deux classes : visage et non visage.

Bibliographie

- [1] R. Min, "Face recognition robust to occlusions," Télécom ParisTech, 2013.
- [2] A. Aissaoui, "Reconnaissance bimodale de visages par fusion de caractéristiques visuelles et de profondeur," Lille 1, 2014.
- [3] A. K. Jain and S. Z. Li, *Handbook of face recognition* vol. 1: Springer, 2005.
- [4] T. G. Kolda and B. W. Bader, "Tensor Decompositions and Applications," *SIAM Review*, vol. 51, pp. 455-500, 2009.
- [5] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "A Multilinear Singular Value Decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, pp. 1253-1278, 2000.
- [6] W.-L. Chao, "Face Recognition," *GICE, National Taiwan University*, 2007.
- [7] M. P. Segundo, C. Queirolo, O. R. P. Bellon, and L. Silva, "Automatic 3D facial segmentation and landmark detection," in *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, 2007, pp. 431-436.
- [8] H. Patil, A. Kothari, and K. Bhurchandi, "3-D face recognition: features, databases, algorithms and challenges," *Artificial Intelligence Review*, vol. 44, pp. 393-441, 2015/10/01 2015.
- [9] A. Chouchane, M. Belahcene, and S. Bourennane, "3D and 2D face recognition using integral projection curves based depth and intensity images," *International Journal of Intelligent Systems Technologies and Applications*, vol. 14, pp. 50-69, 2015.
- [10] Y. Ming, "Rigid-area orthogonal spectral regression for efficient 3D face recognition," *Neurocomputing*, vol. 129, pp. 445-457, 2014.
- [11] Y. Ming, Q. Ruan, and X. Wang, "Efficient 3d face recognition with Gabor patched spectral regression," *Computing and Informatics*, vol. 31, pp. 779--803, 2012.
- [12] A. Mian, M. Bennamoun, and R. Owens, "Face Recognition Using 2D and 3D Multimodal Local Features," in *Advances in Visual Computing*. vol. 4291, G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, A. Nefian, G. Meenakshisundaram, V. Pascucci, J. Zara, J. Molineros, H. Theisel, and T. Malzbender, Eds., ed: Springer Berlin Heidelberg, 2006, pp. 860-870.
- [13] F. Hajati, A. A. Raie, and Y. Gao, "2.5D face recognition using Patch Geodesic Moments," *Pattern Recognition*, vol. 45, pp. 969-982, 2012.
- [14] C. Xu, S. Li, T. Tan, and L. Quan, "Automatic 3D face recognition from depth and intensity Gabor features," *Pattern Recognition*, vol. 42, pp. 1895-1905, 2009.
- [15] M. Belahcene, A. Chouchane, M. Amin Benatia, and M. Halitim, "3D and 2D face recognition based on image segmentation," in *Computational Intelligence for Multimedia Understanding (IWCIM), 2014 International Workshop on*, 2014, pp. 1-5.
- [16] M. Belahcene, A. Chouchane, and N. Mokhtari, "2D and 3D face recognition based on IPC detection and patch of interest regions," in *Connected Vehicles and Expo (ICCVE), 2014 International Conference on*, 2014, pp. 627-628.
- [17] A. Chouchane, M. Belahcene, A. Ouamane, and S. Bourennane, "Multimodal face recognition based on histograms of three local descriptors using score level fusion," in

- Visual Information Processing (EUVIP), 2014 5th European Workshop on*, 2014, pp. 1-6.
- [18] A. Chouchane, M. Belahcene, A. Ouamane, and S. Bourennane, "3D face recognition based on histograms of local descriptors," in *Image Processing Theory, Tools and Applications (IPTA), 2014 4th International Conference on*, 2014, pp. 1-5.
- [19] H. Guesmi, "Identification de personnes par fusion de différentes modalités biométriques," Télécom Bretagne; Université de Rennes 1, 2014.
- [20] R. Heitmeyer, "Biometric identification promises fast and secure processing of airline passengers," *ICAO journal.*, vol. 55, 2000.
- [21] A. Chaari, "Nouvelle approche d'identification dans les bases de données biométriques basée sur une classification non supervisée," Université d'Evry-Val d'Essonne, 2009.
- [22] M. Bennamoun, Y. Guo, and F. Sohel, "Feature selection for 2D and 3D face recognition," *Encyclopedia of electrical and electronics engineering. Book Chapter*, pp. 1-54, 2015.
- [23] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, pp. 399-458, 2003.
- [24] C. Zhang and Z. Zhang, "A survey of recent advances in face detection," Tech. rep., Microsoft Research 2010.
- [25] P. Buysens, "Fusion de différents modes de capture pour la reconnaissance du visage appliquée aux e_transactions," Université de Caen, 2011.
- [26] X. Peng, M. Bennamoun, and A. S. Mian, "A training-free nose tip detection method from face range images," *Pattern Recognition*, vol. 44, pp. 544-558, 2011.
- [27] L. Xiaoguang and A. K. Jain, "Automatic feature extraction for multiview 3D face recognition," in *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, 2006, pp. 585-590.
- [28] C. Xu, T. Tan, Y. Wang, and L. Quan, "Combining local features for robust nose location in 3D facial data," *Pattern Recognition Letters*, vol. 27, pp. 1487-1494, 2006.
- [29] L. Anuar, S. Mashohor, M. Mokhtar, and W. Wan Adnan, "Nose tip region detection in 3D facial model across large pose variation and facial expression," *IJCSI International Journal of Computer Science Issues*, vol. 7, 2010.
- [30] X. Zhang and Y. Gao, "Face recognition across pose: A review," *Pattern Recognition*, vol. 42, pp. 2876-2896, 2009.
- [31] K. Bailly, "Méthodes d'apprentissage pour l'estimation de la pose de la tête dans des images monoculaires," Université Pierre et Marie Curie-Paris VI, 2010.
- [32] L. Xiaoming and C. Tsuhan, "Pose-robust face recognition using geometry assisted probabilistic modeling," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, pp. 502-509 vol. 1.
- [33] R. Gross, J. Shi, and J. F. Cohn, *Quo vadis face recognition?:* Carnegie Mellon University, The Robotics Institute, 2001.
- [34] M. El-Abed, "Evaluation of Biometric Systems," Université de Caen, 2011.
- [35] S. Guerfi Ababsa, "Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D," Evry-Val d'Essonne, 2008.
- [36] L. Ballihi, "Biométrie faciale 3D par apprentissage des caractéristiques géométriques: application à la reconnaissance des visages et à la classification du genre," Lille 1, 2012.

- [37] A. Scheenstra, A. Ruifrok, and R. C. Veltkamp, "A survey of 3D face recognition methods," in *Audio-and Video-Based Biometric Person Authentication*, 2005, pp. 891-899.
- [38] A. S. Mian, M. Bennamoun, and R. Owens, "An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, pp. 1927-1943, 2007.
- [39] P. Viola and M. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, vol. 57, pp. 137-154, 2004/05/01 2004.
- [40] Y. Rodriguez, "Face detection and verification using local binary patterns," Ecole Polytechnique Fédérale de Lausanne, 2006.
- [41] Y. Ming-Hsuan, D. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, pp. 34-58, 2002.
- [42] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, pp. I-511-I-518 vol.1.
- [43] J. Wang, "Spatially Enhanced Local Binary Patterns for Face Detection and Recognition in Mobile Device Applications," University of Toronto, 2013.
- [44] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Li, "Face Detection Based on Multi-Block LBP Representation," in *Advances in Biometrics*. vol. 4642, S.-W. Lee and S. Li, Eds., ed: Springer Berlin Heidelberg, 2007, pp. 11-18.
- [45] T. Mita, T. Kaneko, and O. Hori, "Joint Haar-like features for face detection," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, 2005, pp. 1619-1626 Vol. 2.
- [46] L. Ye, W. YingHui, W. BingBo, and S. LianSheng, "Nose tip detection on three-dimensional faces using pose-invariant differential surface features," *Computer Vision, IET*, vol. 9, pp. 75-84, 2015.
- [47] F. B. ter Haar and R. C. Veltkamp, "A 3D face matching framework," in *Shape Modeling and Applications, 2008. SMI 2008. IEEE International Conference on*, 2008, pp. 103-110.
- [48] L. Xiaoguang, A. K. Jain, and D. Colbry, "Matching 2.5D face scans to 3D models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, pp. 31-43, 2006.
- [49] M. D. Breitenstein, D. Kuettel, T. Weise, L. Van Gool, and H. Pfister, "Real-time face pose estimation from single range images," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1-8.
- [50] K. I. Chang, W. Bowyer, and P. J. Flynn, "Multiple Nose Region Matching for 3D Face Recognition under Varying Facial Expression," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, pp. 1695-1700, 2006.
- [51] N. Pears, T. Heseltine, and M. Romero, "From 3D Point Clouds to Pose-Normalised Depth Maps," *International Journal of Computer Vision*, vol. 89, pp. 152-176, 2010/09/01 2010.

- [52] L. Rong, R. Hu, and Y. Huimin, "Nose detection on 3D face images by depth-based template matching," in *Image and Signal Processing (CISP), 2014 7th International Congress on*, 2014, pp. 302-307.
- [53] A. Slater, H. Yu Hen, and N. Boston, "Multiscale Integral Invariants For Facial Landmark Detection in 2.5D Data," in *Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on*, 2007, pp. 175-178.
- [54] A. Colombo, C. Cusano, and R. Schettini, "3D face detection using curvature analysis," *Pattern recognition*, vol. 39, pp. 444-455, 2006.
- [55] S. Gupta, M. K. Markey, and A. C. Bovik, "Advances and challenges in 3D and 2D+3D human face recognition," *Pattern recognition in biology*, pp. 63-103, 2007.
- [56] N. Erdogmus, "Utilisation des données 3D pour la reconnaissance du visage," Télécom ParisTech, 2012.
- [57] P. Gang, H. Shi, W. Zhaohui, and W. Yueming, "3D Face Recognition using Mapped Depth Images," in *Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, 2005, pp. 175-175.
- [58] T. Huynh, R. Min, and J.-L. Dugelay, "An Efficient LBP-Based Descriptor for Facial Depth Images Applied to Gender Recognition Using RGB-D Face Data," in *Computer Vision - ACCV 2012 Workshops*. vol. 7728, J.-I. Park and J. Kim, Eds., ed: Springer Berlin Heidelberg, 2013, pp. 133-145.
- [59] S. Naveen and R. S. Moni, "Multimodal Face Recognition System Using Spectral Transformation of 2D Texture Feature and Statistical Processing of Face Range Images," *Procedia Computer Science*, vol. 46, pp. 1537-1545, 2015.
- [60] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, pp. 239-256, 1992.
- [61] H. Gao, I. R. Stiefelhagen, and A. Waibel, "Local Appearance-based 3D Face Recognition," 2006.
- [62] G. Medioni and R. Waupotitsch, "Face modeling and recognition in 3-D," in *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on*, 2003, pp. 232-233.
- [63] B. B. Amor, K. Ouji, M. Ardabilian, and L. Chen, "3D Face recognition by ICP-based shape matching."
- [64] A. B. Moreno, A. Sánchez, J. F. Vélez, and F. J. Díaz, "Face recognition using 3D surface-extracted descriptors," in *Irish Machine Vision and Image Processing Conference*, 2003.
- [65] D. J. Bardsley and L. Bai, "3D surface reconstruction and recognition," in *Defense and Security Symposium*, 2007, pp. 653906-653906-11.
- [66] V. Blanz, S. Romdhani, and T. Vetter, "Face identification across different poses and illuminations with a 3D morphable model," in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, 2002, pp. 192-197.
- [67] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, pp. 1063-1074, 2003.

- [68] 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.
- [69] Y. Dong, L. Zhen, and S. Z. Li, "Towards Pose Robust Face Recognition," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 3539-3545.
- [70] Z. Dianle, D. Petrovska-Delacretaz, and B. Dorizzi, "3D Active Shape Model for Automatic Facial Landmark Location Trained with Automatically Generated Landmark Points," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, 2010, pp. 3801-3805.
- [71] S. K. Zhou and R. Chellappa, "Illuminating light field: image-based face recognition across illuminations and poses," in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, 2004, pp. 229-234.
- [72] C. Heshner, A. Srivastava, and G. Erlebacher, "A novel technique for face recognition using range imaging," in *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*, 2003, pp. 201-204 vol.2.
- [73] K. Chang, K. Bowyer, and P. Flynn, "Face recognition using 2D and 3D facial data," in *ACM Workshop on Multimodal User Authentication*, 2003, pp. 25-32.
- [74] T. Papatheodorou and D. Rueckert, "Evaluation of 3D Face Recognition Using Registration and PCA," in *Audio- and Video-Based Biometric Person Authentication*. vol. 3546, T. Kanade, A. Jain, and N. Ratha, Eds., ed: Springer Berlin Heidelberg, 2005, pp. 997-1009.
- [75] F. Tsalakanidou, S. Malassiotis, and M. G. Strintzis, "Integration of 2D and 3D images for enhanced face authentication," in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, 2004, pp. 266-271.
- [76] B. Achermann, X. Jiang, and H. Bunke, "Face recognition using range images," in *Virtual Systems and MultiMedia, 1997. VSMM'97. Proceedings., International Conference on*, 1997, pp. 129-136.
- [77] F. R. Al-Osaimi, M. Bennamoun, and A. Mian, "Integration of local and global geometrical cues for 3D face recognition," *Pattern Recognition*, vol. 41, pp. 1030-1040, 2008.
- [78] T. Russ, C. Boehnen, and T. Peters, "3D face recognition using 3D alignment for PCA," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, 2006, pp. 1391-1398.
- [79] T. Heseltine, N. Pears, and J. Austin, "Three-dimensional face recognition: A fishersurface approach," in *Image Analysis and Recognition*, ed: Springer, 2004, pp. 684-691.
- [80] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, pp. 711-720, 1997.
- [81] L. Spreeuwers, "Fast and accurate 3d face recognition," *International Journal of Computer Vision*, vol. 93, pp. 389-414, 2011.
- [82] S. Gupta, M. K. Markey, and A. C. Bovik, "Anthropometric 3D face recognition," *International journal of computer vision*, vol. 90, pp. 331-349, 2010.

- [83] S. Gupta, M. K. Markey, J. Aggarwal, and A. C. Bovik, "Three dimensional face recognition based on geodesic and euclidean distances," in *Electronic Imaging 2007*, 2007, pp. 64990D-64990D-11.
- [84] F. Sohel, M. Bennamoun, and M. Hahn, "Spatial shape error concealment utilising image texture," in *Industrial Electronics and Applications (ICIEA), 2011 6th IEEE Conference on*, 2011, pp. 265-270.
- [85] K. Jongsun, C. Jongmoo, Y. Juneho, and M. Turk, "Effective representation using ICA for face recognition robust to local distortion and partial occlusion," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, pp. 1977-1981, 2005.
- [86] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi, "Recent advances in visual and infrared face recognition—a review," *Computer Vision and Image Understanding*, vol. 97, pp. 103-135, 2005.
- [87] P. Liu, Y. Wang, D. Huang, Z. Zhang, and L. Chen, "Learning the spherical harmonic features for 3-D face recognition," *Image Processing, IEEE Transactions on*, vol. 22, pp. 914-925, 2013.
- [88] H.-S. Wong, K. K. Cheung, and H. H. Ip, "3D head model classification by evolutionary optimization of the Extended Gaussian Image representation," *Pattern Recognition*, vol. 37, pp. 2307-2322, 2004.
- [89] C. Beumier and M. Acheroy, "Automatic 3D face authentication," *Image and Vision Computing*, vol. 18, pp. 315-321, 2000.
- [90] C. Beumier and M. Acheroy, "Face verification from 3D and grey level clues," *Pattern Recognition Letters*, vol. 22, pp. 1321-1329, 2001.
- [91] H. Drira, B. Ben Amor, A. Srivastava, and M. Daoudi, "A Riemannian analysis of 3D nose shapes for partial human biometrics," in *Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 2050-2057.
- [92] Y. Lee, H. Song, U. Yang, H. Shin, and K. Sohn, "Local Feature Based 3D Face Recognition," in *Audio- and Video-Based Biometric Person Authentication*. vol. 3546, T. Kanade, A. Jain, and N. Ratha, Eds., ed: Springer Berlin Heidelberg, 2005, pp. 909-918.
- [93] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," *International Journal of Computer Vision*, vol. 64, pp. 5-30, 2005.
- [94] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Expression-Invariant Representations of Faces," *Image Processing, IEEE Transactions on*, vol. 16, pp. 188-197, 2007.
- [95] H. Di, M. Ardabilian, W. Yunhong, and C. Liming, "A novel geometric facial representation based on multi-scale extended local binary patterns," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, 2011, pp. 1-7.
- [96] M. Belahcene, "Authentification et identification en biométrie," Université Mohamed Khider de Biskra, 2013.
- [97] A. OUAMANE, "Reconnaissance Biométrique par Fusion Multimodale du Visage 2D et 3D," Université Mohamed Khider-Biskra, 2015.
- [98] S. Li, C. Zhao, M. Ao, and Z. Lei, "Learning to Fuse 3D+2D Based Face Recognition at Both Feature and Decision Levels," in *Analysis and Modelling of Faces and Gestures*.

- vol. 3723, W. Zhao, S. Gong, and X. Tang, Eds., ed: Springer Berlin Heidelberg, 2005, pp. 44-54.
- [99] G. García Mateos and C. Vicente Chicote, "Face Detection on Still Images Using HIT Maps," in *Audio- and Video-Based Biometric Person Authentication*. vol. 2091, J. Bigun and F. Smeraldi, Eds., ed: Springer Berlin Heidelberg, 2001, pp. 102-107.
- [100] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, pp. 23-27, 1975.
- [101] P. Bagchi, D. Bhattacharjee, M. Nasipuri, and D. K. Basu, "A novel approach for nose tip detection using smoothing by weighted median filtering applied to 3D face images in variant poses," in *Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on*, 2012, pp. 272-277.
- [102] A. A. Ross and R. Govindarajan, "Feature level fusion of hand and face biometrics," in *Defense and Security*, 2005, pp. 196-204.
- [103] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, 1991, pp. 586-591.
- [104] Y. Ma, S. Lao, E. Takikawa, and M. Kawade, "Discriminant analysis in correlation similarity measure space," presented at the Proceedings of the 24th international conference on Machine learning, Corvallis, Oregon, USA, 2007.
- [105] M. Paci, L. Nanni, A. Lahti, K. Aalto-Setälä, J. Hyttinen, and S. Severi, "Non-binary coding for texture descriptors in sub-cellular and stem cell image classification," *Current Bioinformatics*, vol. 8, pp. 208-219, 2013.
- [106] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273-297, 1995.
- [107] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144-152.
- [108] S. Meshgini, A. Aghagolzadeh, and H. Seyedarabi, "Face recognition using Gabor-based direct linear discriminant analysis and support vector machine," *Computers & Electrical Engineering*, vol. 39, pp. 727-745, 2013.
- [109] Y. Lei, M. Bennamoun, and A. A. El-Sallam, "An efficient 3D face recognition approach based on the fusion of novel local low-level features," *Pattern Recognition*, vol. 46, pp. 24-37, 2013.
- [110] H. Chih-Wei and L. Chih-Jen, "A comparison of methods for multiclass support vector machines," *Neural Networks, IEEE Transactions on*, vol. 13, pp. 415-425, 2002.
- [111] L. Zhao, Y. Song, Y. Zhu, C. Zhang, and Y. Zheng, "Face recognition based on multi-class SVM," in *Control and Decision Conference, 2009. CCDC '09. Chinese*, 2009, pp. 5871-5873.
- [112] A. B. Moreno, A. Sanchez, J. Velez, and J. Diaz, "Face recognition using 3D local geometrical features: PCA vs. SVM," in *Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on*, 2005, pp. 185-190.

- [113] L. Xiaoxing, J. Tao, and Z. Hao, "Expression-insensitive 3D face recognition using sparse representation," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 2575-2582.
- [114] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D Face Recognition under Expressions, Occlusions, and Pose Variations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, pp. 2270-2283, 2013.
- [115] M. H. Mahoor and M. Abdel-Mottaleb, "Face recognition based on 3D ridge images obtained from range data," *Pattern Recognition*, vol. 42, pp. 445-451, 2009.
- [116] M. H. Mousavi, K. Faez, and A. Asghari, "Three Dimensional Face Recognition Using SVM Classifier," in *Computer and Information Science, 2008. ICIS 08. Seventh IEEE/ACIS International Conference on*, 2008, pp. 208-213.
- [117] M. Yue, R. Qiuqi, W. Xueqiao, and M. Meiru, "Robust 3D face recognition using learn correlative features," in *Signal Processing (ICSP), 2010 IEEE 10th International Conference on*, 2010, pp. 1382-1385.
- [118] J. Zhou, Y. Li, and J. Wang, "2D&3D-ComFusFace: 2D and 3D Face Recognition by Scalable Fusion of Common Features," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 12, p. 30, 2012.
- [119] X. Zhou, H. Seibert, C. Busch, and W. Funk, "A 3d face recognition algorithm using histogram-based features," in *Proceedings of the 1st Eurographics conference on 3D Object Retrieval*, 2008, pp. 65-71.
- [120] H.-D. Liu, M. Yang, Y. Gao, and C. Cui, "Local histogram specification for face recognition under varying lighting conditions," *Image and Vision Computing*, vol. 32, pp. 335-347, 2014.
- [121] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based methods in the wild," in *Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition*, 2008.
- [122] A. Aissaoui, J. Martinet, and C. Djeraba, "DLBP: A novel descriptor for depth image based face recognition," in *Image Processing (ICIP), 2014 IEEE International Conference on*, 2014, pp. 298-302.
- [123] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila, "Recognition of blurred faces using Local Phase Quantization," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, 2008, pp. 1-4.
- [124] C. Chi Ho, M. A. Tahir, J. Kittler, Pietika, x, and M. inen, "Multiscale Local Phase Quantization for Robust Component-Based Face Recognition Using Kernel Fusion of Multiple Descriptors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, pp. 1164-1177, 2013.
- [125] Y. Baohua, C. Honggen, and C. Jiuliang, "Combining Local Binary Pattern and Local Phase Quantization for Face Recognition," in *Biometrics and Security Technologies (ISBAST), 2012 International Symposium on*, 2012, pp. 51-53.
- [126] S. Berretti, A. D. Bimbo, and P. Pala, "3D face recognition using isogeodesic stripes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, pp. 2162-2177, 2010.
- [127] J. Kannala and E. Rahtu, "Bsif: Binarized statistical image features," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 1363-1366.

- [128] C. Deng, H. Xiaofei, H. Jiawei, and Z. Hong-Jiang, "Orthogonal Laplacianfaces for Face Recognition," *Image Processing, IEEE Transactions on*, vol. 15, pp. 3608-3614, 2006.
- [129] W. Li, L. Ruifeng, W. Ke, and C. Chuqing, "OLPP-based Gabor feature dimensionality reduction for facial expression recognition," in *Information and Automation (ICIA), 2014 IEEE International Conference on*, 2014, pp. 455-460.
- [130] M. Bereta, W. Pedrycz, and M. Reformat, "Local descriptors and similarity measures for frontal face recognition: A comparative analysis," *Journal of Visual Communication and Image Representation*, vol. 24, pp. 1213-1231, 2013.
- [131] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face Recognition with Local Binary Patterns," in *Computer Vision - ECCV 2004*. vol. 3021, T. Pajdla and J. Matas, Eds., ed: Springer Berlin Heidelberg, 2004, pp. 469-481.
- [132] C. H. Chan, M. A. Tahir, J. Kittler, and M. Pietikainen, "Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, pp. 1164-1177, 2013.
- [133] S. Li, D. Gong, and Y. Yuan, "Face recognition using Weber local descriptors," *Neurocomputing*, vol. 122, pp. 272-283, 2013.
- [134] Z. Cui, S. Shan, R. Wang, L. Zhang, and X. Chen, "Sparsely encoded local descriptor for face verification," *Neurocomputing*, vol. 147, pp. 403-411, 2015.
- [135] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, pp. 2037-2041, 2006.
- [136] C. H. Chan, J. Kittler, N. Poh, T. Ahonen, and M. Pietikainen, "(Multiscale) Local Phase Quantisation histogram discriminant analysis with score normalisation for robust face recognition," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, 2009, pp. 633-640.
- [137] Z. Lei, T. Ahonen, M. Pietikäinen, and S. Z. Li, "Local frequency descriptor for low-resolution face recognition," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, 2011, pp. 161-166.
- [138] V. Ojansivu and J. Heikkilä, "Blur Insensitive Texture Classification Using Local Phase Quantization," in *Image and Signal Processing*. vol. 5099, A. Elmoataz, O. Lezoray, F. Nouboud, and D. Mammass, Eds., ed: Springer Berlin Heidelberg, 2008, pp. 236-243.
- [139] C. Fiche, "Repousser les limites de l'identification faciale en contexte de vidéo-surveillance," Grenoble, 2012.
- [140] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, pp. 971-987, 2002.
- [141] A. Hadid, J. Ylioinas, and M. B. Lopez, "Face and texture analysis using local descriptors: A comparative analysis," in *Image Processing Theory, Tools and Applications (IPTA), 2014 4th International Conference on*, 2014, pp. 1-4.
- [142] L. Liu, Y. Long, P. W. Fieguth, S. Lao, and G. Zhao, "BRINT: Binary rotation invariant and noise tolerant texture classification," *Image Processing, IEEE Transactions on*, vol. 23, pp. 3071-3084, 2014.

- [143] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of Interest Regions with Center-Symmetric Local Binary Patterns," in *Computer Vision, Graphics and Image Processing*. vol. 4338, P. Kalra and S. Peleg, Eds., ed: Springer Berlin Heidelberg, 2006, pp. 58-69.
- [144] A. Rattani, C. Chen, and A. Ross, "Evaluation of Texture Descriptors for Automated Gender Estimation from Fingerprints," in *Computer Vision - ECCV 2014 Workshops*. vol. 8926, L. Agapito, M. M. Bronstein, and C. Rother, Eds., ed: Springer International Publishing, 2015, pp. 764-777.
- [145] A. Hyvärinen, J. Hurri, and P. O. Hoyer, *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision* vol. 39: Springer Science & Business Media, 2009.
- [146] M. Roschani, "Evaluation of Local Descriptors on the Labeled Faces in the Wild Dataset," 2009.
- [147] M. Belkin and P. Niyogi, "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering," in *NIPS*, 2001, pp. 585-591.
- [148] W. Xueqiao, R. Qiuqi, and M. Yue, "3D Face recognition using Corresponding Point Direction Measure and depth local features," in *Signal Processing (ICSP), 2010 IEEE 10th International Conference on*, 2010, pp. 86-89.
- [149] L. Yong-An, S. Yong-Jun, Z. Gui-Dong, Y. Taohong, X. Xiu-Ji, and X. Hua-Long, "An Efficient 3D Face Recognition Method Using Geometric Features," in *Intelligent Systems and Applications (ISA), 2010 2nd International Workshop on*, 2010, pp. 1-4.
- [150] A. Ouamane, M. Belahcene, A. Benakcha, S. Bourennane, and A. Taleb-Ahmed, "Robust multimodal 2D and 3D face authentication using local feature fusion," *Signal, Image and Video Processing*, pp. 1-9, 2014/11/16 2014.
- [151] G. Fanelli, T. Weise, J. Gall, and L. V. Gool, "Real time head pose estimation from consumer depth cameras," presented at the Proceedings of the 33rd international conference on Pattern recognition, Frankfurt, Germany, 2011.
- [152] G. Passalis, P. Perakis, T. Theoharis, and I. A. Kakadiaris, "Using Facial Symmetry to Handle Pose Variations in Real-World 3D Face Recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, pp. 1938-1951, 2011.
- [153] H. Dibeklioglu, A. A. Salah, and L. Akarun, "3D Facial Landmarking under Expression, Pose, and Occlusion Variations," in *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*, 2008, pp. 1-6.
- [154] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "Rotated Profile Signatures for robust 3D feature detection," in *Automatic Face & Gesture Recognition, 2008. FG '08. 8th IEEE International Conference on*, 2008, pp. 1-7.
- [155] A. Rajwade and M. D. Levine, "Facial pose from 3D data," *Image and Vision Computing*, vol. 24, pp. 849-856, 2006.
- [156] W. Zhang, D. Huang, Y. Wang, and L. Chen, "3D Aided Face Recognition across Pose Variations," in *Biometric Recognition*. vol. 7701, W.-S. Zheng, Z. Sun, Y. Wang, X. Chen, P. Yuen, and J. Lai, Eds., ed: Springer Berlin Heidelberg, 2012, pp. 58-66.
- [157] M. A. O. Vasilescu, "A Multilinear (Tensor) Algebraic Framework for Computer Graphics, Computer Vision, and Machine Learning," University of Toronto, 2009.

- [158] D. Letexier, "Filtrages tensoriels adaptatifs pour la restauration d'images multidimensionnelles," Aix-Marseille 3, 2009.
- [159] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "A Multilinear Singular Value Decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, pp. 1253-1278, 2000.
- [160] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, pp. 279-311, 1966.