



Faculté des sciences exactes et des sciences de la nature et de la vie
Département d'informatique

THÈSE

Pour l'obtention du grade de

DOCTEUR 3^{ème} CYCLE EN INFORMATIQUE

Option : Image et Vie Artificielle

Titre

Reconnaissance tridimensionnelle du visage

Par

SAAD Narimen

Soutenue le :

Devant le jury composé de

Président :	Dr Foudil CHERIF	Professeur	Université Med KHIDER, Biskra.
Rapporteur :	Dr NouEddine DJEDI	Professeur	Université Med KHIDER, Biskra.
Examineur :	Dr Zineddine BAARIR	Professeur	Université Med KHIDER, Biskra.
	Dr A/Wahab MOUSSAOUI	Professeur	Université Ferhat ABBAS, Sétif 1,
	Dr Med Lamine KHERFI,	MCA	Université Kasdi MERBAH, Ouargla.

Biskra, Algérie.

Dédicace

À

mes parents,

mon mari,

mes frères,

mes enseignants,

et mes amis,

*qui ont dû me supporter, pour leur soutien et encouragement ainsi que pour leur
grande aide dans la rédaction de ma thèse,*

je dédie ce modeste travail.

Remerciements

Je tiens premièrement à me prosterner remerciant Allah le Tout-Puissant de m'avoir donnée le courage et la patience pour terminer ce travail.

Je remercie ensuite mon encadreur monsieur le Professeur NourEddine DJEDI, de m'avoir fait l'honneur de diriger ce travail, de ses précieux conseils, de sa disponibilité, de sa confiance scientifique en moi et de ses discussions très riches qui m'ont permis de découvrir et maîtriser le domaine de la recherche scientifique.

Mes remerciements vont également aux membres de jury : Dr Foudil CHERIF, Dr Zineddine BAARIR, Dr A/Wahab MOUSSAOUI et Dr Mohamed Lamine KHERFI pour m'avoir honorée en acceptant d'évaluer mon travail de thèse.

Un grand merci va également à l'ensemble des membres du laboratoire LESIA avec qui j'ai partagé bien plus qu'un lieu de travail.

Je désire aussi remercier tous mes amis qui ont dû me supporter et mes collègues.

Résumé

De nos jours, on parle de plus en plus de l'insécurité dans divers secteurs ainsi que celle liée à l'exploitation des moyens informatiques. Afin de renforcer la sécurité dans tous ces domaines, devenue aujourd'hui un enjeu majeur sous l'effet de la mondialisation et de la crise économique, la biométrie se présente comme une technologie potentiellement puissante. Un système biométrique est essentiellement un système de reconnaissance qui utilise les données biométriques d'un individu. La problématique générale de cette thèse est de développer une technique de reconnaissance 3D de visages. La reconnaissance faciale 3D est une alternative prometteuse à la résolution de problèmes de reconnaissance 2D. Pour cette raison, deux contributions ont été développées. La première contribution dans laquelle nous discutons spécifiquement les difficultés majeures pour proposer et tester une nouvelle solution de reconnaissance faciale 3D capable de réaliser le processus de reconnaissance, dans les cas où l'analyse concerne une partie du visage seulement. Avec l'approche proposée, les caractéristiques distinctives du visage sont capturées en extrayant d'abord les points-clés SIFT sur le visage de l'analyse et en mesurant comment le visage change le long des profils construits entre paires de points-clés, ensuite nous appliquons l'opérateur SIFT sur les images $LBP_{P,R}$ séparément. Suite au travail de Faltemier et al. puis à celui de Tang et al., nous pouvons mieux détecter un certain nombre de points clés en utilisant SIFT sur les images $LBP_{P,R}$, que d'utiliser SIFT sur les images originales. La contribution est testée en utilisant l'ensemble des données Face Recognition Grand Challenge FRGC v1.0. Enfin, nous effectuons une classification basée sur le technique SVM. La deuxième contribution consiste à présenter une technique de reconnaissance des expressions faciales avec images 3D. Ce système utilise un flux composé simultanément d'une image couleur (2D) et d'une image de profondeur (3D) capturée par le capteur Microsoft Kinect. Notre système détecte et identifie l'expression du visage capturé, puis reconnaît les sept expressions développées par Paul Ekman en utilisant la méthode MB-LBP. Nous avons introduit un nouveau jeu de données correspondant à 10 sujets humains pour l'étape d'apprentissage avec une classification en utilisant la méthode SVM. L'implémentation de notre solution est réalisée en C++ et exploite de plus la bibliothèque OpenCV.

Les mots clés : Biométrie, reconnaissance faciale 3D, expressions faciales, Kinect, MB-LBP, SVM, reconnaissance faciale, base de données FRGC V1.0.

ملخص

في الوقت الحاضر، نتحدث أكثر فأكثر عن انعدام الأمن في مختلف القطاعات وكذلك في مجال تكنولوجيا المعلومات. ومن أجل تعزيز الأمن في هذه الميادين المختلفة، التي أصبحت اليوم معضلة عالمية نتيجة للأزمة الاقتصادية، يعد النظام البيومتري تكنولوجيا قوية لحلحلة هذه المشاكل. النظام البيومتري هو في الأساس نظام الاعتراف الذي يستخدم البيانات البيومترية للفرد. وتمثل الإشكالية الشاملة من هذه الدراسة في تطوير تقنية التعرف على الوجه الثلاثي الأبعاد، التعرف على الوجه الثلاثي الأبعاد هو بديل واعد لحل مشاكل التعرف على الوجه ثنائي الأبعاد.

وفي هذا الصدد، تم تطوير نموذجين:

- النموذج الأول ناقش فيه على وجه التحديد الصعوبات الرئيسية لاقتراح واختبار جديد لحل التعرف على الوجه الثلاثي الأبعاد وتكون قادرة على تحقيق موضوع الاعتراف، في حالة تحليل جزء فقط من الوجه. مع النهج المقترح، يتم التقاط السمات المميزة للوجه من خلال استخراج أول النقاط الرئيسية باستخدام تقنية SIFT على وجه التحليل وقياس كيفية تغير الوجه على طول التشكيلات التي شيدت بين أزواج من النقاط الرئيسية، ثم تطبيق تقنية SIFT على الصور $LBP_{P,R}$ بشكل منفصل. باتباع أعمال Faltemier et al. وكذلك Tang et al. ، يمكننا الكشف بشكل أفضل عن عدد من النقاط الرئيسية باستخدام SIFT على الصور $LBP_{P,R}$ ، من استخدام SIFT على الصور الأصلية. يتم اختبار النموذج باستخدام قاعدة البيانات للتعرف على الوجه FRGC V1.0. وأخيرا، نقوم بإجراء تصنيف على أساس عملية SVM.
- النموذج الثاني يتم فيه تقديم نهجنا المقترح للتعرف على تعبيرات الوجه مع صورة ثلاثية الأبعاد. يستخدم هذا النظام تيار يتكون في وقت واحد من صورة ملونة (D2) وصورة العمق (D3) التي تم التقاطها بواسطة جهاز الكاميرا المطور من طرف ميكروسوفت المتمثلة في كينكت (Kinect). يكتشف نظامنا ويحدد تعبيرات الوجه الملتقطة، بحيث يعتمد على تعبيرات الوجه السبعة التي وضعها "بول إيكمان (Paul Ekman)" وذلك باستخدام طريقة MB_LBP. قدمنا مجموعة بيانات جديدة توافق 10 أفراد من أجل خطوة التمرين مع التصنيف باستخدام طريقة SVM. يتم تنفيذ لبرمجة باستخدام البرمجة بواسطة ++C وكذلك مكتبة OpenCV.

الكلمات المفتاحية: البيومترية، التعرف على الوجه الثلاثي الأبعاد، تعبيرات الوجه، كينكت (Kinect)، MB_LBP ، SVM، التعرف على الوجه ، قاعدة البيانات FRGC V1.0.

Abstract

Nowadays we are talking more and more about insecurity in various sectors as well as IT resources. In order to enhance security in these various fields, which has become an international concern today as a result of the economic crisis; biometrics represent a potentially powerful technology. A biometric system is essentially a recognition system that uses the biometric data of an individual. The general problematic of this thesis is to develop a 3D face recognition technique. 3D Face Recognition is a promising alternative to solving problems of 2D recognition, in this reason two contributions have been developed. The first contribution, we specifically discuss major difficulties further to propose and test a novel solution of 3D face recognition that is significantly capable to perform the recognition subject, in cases where the analysis of only a part of the face. With the proposed approach, the distinctive features of the face are captured by first extracting SIFT key points on the face of analysis and measure how the face changes along profiles built between pairs of key points; second, we applied the operator SIFT on $LBP_{P, or}$ images, separately. Following the work of Faltemier and al. then Tang and al., we can better detect a number of key points by using SIFT on $LBP_{P,R}$ images, than using SIFT on the original images. The contribution is tested using the whole of the Face Recognition Grand Challenge FRGC v1.0 data. Finally, we perform a classification based on SVM process. The Second contribution is to present our proposed approach to facial expression recognition with 3D image. This system uses a flow composed simultaneously of a color image (2D) and a depth image (3D) captured by the Microsoft Kinect sensor. Our system detects, and identifies the expression of the captured face; then recognize the seven expressions developed by Paul Ekman using the MB-LBP method. We have introduced a new dataset corresponding to 10 human subjects for learning step with a classification using SVM method. The implementation of our solution produces in C++ using the OpenCV library.

Keywords: Biometrics, 3D Face recognition, facial expressions, Kinect, MB-LBP, SVM, Face recognition, FRGC V1.0 database.

Publications et communications

Publications

1. Chapter Book: N. Saad, N. Djedi. Recognition of 3D Faces with Missing Parts based on SIFT and LBP methods. Security and Privacy, Signal Processing for Security Technologies, in: Springer International Publishing Switzerland 2017 Book-Springer. R.Jiang et al. (Eds.), DOI 10.1007/978-3-319-47301-7_12, pages 273-297, 2017.
2. N. Saad, N. Djedi, 3D Face recognition related with facial expressions based on MB-LBP method, Courrier du Savoir, Vol. 25, pages 93-102, 2018.

Communications

1. N. Saad, N. Djedi. 2ème Workshop. Reconnaissance tridimensionnelle du visage 3D Image Graphique et Vie Artificielle (IGVA -2014), 8-9 et 10 Juin 2014. Université Mohamed Khider, Biskra, Algerie.
2. N. Saad, N. Djedi. 3D face recognition based on the profiles Networks. 3rd International Conference on Signal, Image, Vision and their Applications (SIVA'15), ID 45, 23-25 November 2015, Guelma, Algerie.
3. N. Saad, N. Djedi. 3rd International Conference on Automation, Control, Engineering and Computer (ACECS-2016), pages 127-133, IPCO 2016, volume 18 issue 2016, 20-22 Mars 2016, Hammamet, Tunisie.

Table des matières

TABLE DES MATIERES	I
TABLE DES FIGURES.....	IV
LISTE DES TABLEAUX.....	VI
INTRODUCTION GENERALE	1
1. CHAPITRE1 : INTRODUCTION A LA BIOMETRIE.....	4
1.1 INTRODUCTION	4
1.1.1. Définition de la biométrie.....	4
1.1.2. Panorama des différentes techniques biométries	4
1.1.3. Les systèmes biométriques	7
1.1.4. Les modules d'un système biométrique	10
1.1.5. Pourquoi la reconnaissance de visage	11
1.1.6. Systèmes biométriques basés sur la reconnaissance de visage	12
1.1.7. Mode de reconnaissance	14
1.2 LA RECONNAISSANCE DE VISAGE	14
1.2.1. Processus de reconnaissance des visages	14
1.2.2. Principales difficultés de la reconnaissance de visage	15
1.2.3. Opportunités de reconnaissance de visage 3D.....	17
1.2.4. Reconnaissance de visage 3D	18
1.2.5. Base de données 3D de visage	20
1.3 CONCLUSION.....	21
2. CHAPITRE2 : LA RECONNAISSANCE FACIALE ET DIFFERENTES TECHNIQUES	22
2.1 INTRODUCTION	22
2.2 RECONNAISSANCE DE VISAGE 2D.....	22
2.2.1. Approches globales.....	24
2.2.2. Approches locales	28
2.2.3. Autres approches	33
2.2.4. Discussion	34
2.3 RECONNAISSANCE DE VISAGE 3D.....	34
2.3.1. Approches globales	35
2.3.2. Approches locales	37
2.3.3. Autres approches	40
2.3.4. Discussion	40
2.4 RECONNAISSANCE DE L'EXPRESSION FACIALE.....	41
2.4.1. Qu'est ce qu'une expression faciale	41
2.5 RECONNAISSANCE DE VISAGE 2D+3D.....	45
2.6 CONCLUSION.....	46

3. CHAPITRE3 : NOTRE SYSTEME POUR LA RECONNAISSANCE DE VISAGE EN 2D.....	47
3.1 INTRODUCTION	47
3.2 VERIFICATION UNI-MODALE DE VISAGE 2D	47
3.2.1. Acquisition de données.....	48
3.2.2. Prétraitement d'image	49
3.2.3. Détection.....	51
3.2.4. Caractérisation	51
3.2.5. Apprentissage.....	54
3.2.6. Reconnaissance	56
3.3 VERIFICATION MULTI-ALGORITHMES DE VISAGE 2D	56
3.3.1. Limitations des systèmes biométriques monomodaux.....	56
3.3.2. Les différentes multi-modalités possibles	57
3.3.3. État de l'art de la fusion de scores.....	58
3.3.4. La fusion de scores	59
3.3.5. Description de la reconnaissance d'expressions faciales	60
3.4 CONCLUSION.....	63
4. CHAPITRE4 : NOTRE METHODE POUR LAVERIFICATION DES VISAGES 3D.....	64
4.1 INTRODUCTION	64
4.2 PREMIERE CONTRIBUTION : LA RECONNAISSANCE DE VISAGE EN 3D LIEES AUX EXPRESSIONS FACIALES BASEE SUR LA METHODE MB_LBP	64
4.2.1. Description des étapes de l'approche proposée.....	65
4.2.2. Présentation du Kinect.....	72
4.3 RECONSTRUCTION ET ANIMATION3D DES EXPRESSIONS FACIALES	74
4.3.1. Description de la reconstruction 3D.....	74
4.3.2. Vision stéréoscopique	74
4.3.3. Description de l'animation 3D.....	75
4.3.4. Techniques existantes	75
4.4 DEUXIEME CONTRIBUTION : LA RECONNAISSANCE DE VISAGE 3D BASE SUR LA FUSION SIFT+LBP EN UTILISANT LA BASE DE DONNEES FRGC v1.0.....	77
4.4.1. Base de données FRGC V1.0 3D	77
4.4.2. Description des étapes de l'approche proposée.....	77
4.4.3. Scores Fusion	85
4.5 CONCLUSION.....	86
5. CHAPITRE5 : IMPLEMENTATION, RESULTATS ET COMPARAISON.....	87
5.1 INTRODUCTION	87
5.2 CONTRIBUTION I : IMPLEMENTATIONS, RESULTATS ET COMPARAISON	87
5.2.1. Système automatique 'Temps réel'	87
5.2.2. Résultats : Présentation des interfaces de notre Système	96

5.2.3. Comparaison de notre système (AREFK) avec d'autres travaux	100
5.3 CONTRIBUTION II : IMPLEMENTATIONS, RESULTATS ET COMPARAISON.....	102
5.3.1. Base de données FRGC V1.0 3D	102
5.3.2. Résultats	103
5.3.3. Comparaison des algorithmes de réduction de l'espace de données	109
5.3.4. Comparaison avec d'autres méthodes et d'autres travaux	110
5.4 CONCLUSION.....	112
CONCLUSION GENERALE ET PERSPECTIVES	113
REFERENCES BIBLIOGRAPHIQUES	115

Liste des figures

FIGURE 1.1 : QUELQUES MODALITES BIOMETRIQUES.....	5
FIGURE 1.2 : QUATRE TYPES DE FORMATS D'ENREGISTREMENT DU VISAGE 3D	6
FIGURE 1.3 : ILLUSTRATION DU FRR ET DU FAR.	9
FIGURE 1.4 : COURBE ROC	9
FIGURE 1.5 : COURBES CMC DU CSU SYSTEM 5.0.	10
FIGURE 1.6 : PRINCIPAUX MODULES D'UN SYSTEME BIOMETRIQUE.....	11
FIGURE 1.7 : LE SCHEMA GENERAL D'UN SYSTEME DE RECONNAISSANCE DE VISAGE	13
FIGURE 1.8 : EXEMPLE DE VARIATION D'ECLAIRAGE.....	15
FIGURE 1.9 : EXEMPLES DE VARIATION DE POSES.	16
FIGURE 1.10 : EXEMPLES DE VARIATION D'EXPRESSIONS.	16
FIGURE 1.11 : EXEMPLES D'OCCLUSION DU VISAGE.	17
FIGURE 1.12 : PROCESSUS DE RECONNAISSANCE FACIALE 3D	18
FIGURE 1.13 : LA PROJECTION D'UNE LUMIERE STRUCTUREE SUR UN VISAGE	19
FIGURE 1.14 : COMPARAISON DE LA STEREO PHOTOMETRIQUE ET PHOTOMETRIQUE MULTI SPECTRALE.....	19
FIGURE 1.15 : SYSTEMES DE CAPTURE.	20
FIGURE 2.1 : CONVERSION D'IMAGE EN VECTEUR COLONNE.	25
FIGURE 2.2 : CALCUL DE LA MATRICE DE POIDS.	25
FIGURE 2.3 : EXEMPLES DES 5 PREMIERS VISAGES PROPRES CONSTRUITS AVEC L'APPROCHE ACP.....	26
FIGURE 2.4 : PROJECTION PAR LDA AVEC SEPARATION ENTRE LES CLASSES.	27
FIGURE 2.5 : CONSTRUCTION D'UN DESCRIPTEUR SIFT.....	30
FIGURE 2.6 : LES CARACTERISTIQUES MB_LBP PROPOSEES PAR ZHANG ET AL	31
FIGURE 2.7 : LA CHAINE D'EXECUTION DE LA METHODE « VIOLA & JONES »	32
FIGURE 2.8 : EXTRAIT DE MUSCLES DE VISAGE.....	42
FIGURE 3.1 : SCHEMA GENERAL DE NOTRE APPROCHE PROPOSEE.....	48
FIGURE 3.2 : DETAILS DE L'ETAPE « ACQUISITION »	48
FIGURE 3.3 : IMAGE DE VISAGE (A) AVANT ET (B) APRES DECOUPAGE.	49
FIGURE 3.4 : IMAGE DE VISAGE (A) AVANT (B) APRES DECIMATION.....	49
FIGURE 3.5 : DIFFERENTS SYSTEMES COLORIMETRIQUES	50
FIGURE 3.6 : LES EQUATIONS DES TRANSFORMATIONS DES ESPACES RVB VERS TSV.	50
FIGURE 3.7 : IMAGE DECOUPEE ET DECIMEE	51
FIGURE 3.8 : L'OPERATEUR LBP	52
FIGURE 3.9 : TROIS VOISINAGES POUR DES R ET P DIFFERENTS	52
FIGURE 3.10 : UN EXEMPLE DE CODE UNIFORMES POUR P=8.....	54
FIGURE 3.11 : MONTRE EXEMPLE DE LA METHODE MB_LBP POUR UN VISAGE.....	54
FIGURE 3.12 : SEPARATION LINEAIRE DANS UN ESPACE A DEUX DIMENSIONS	55
FIGURE 3.13 : LES DIFFERENTS SYSTEMES MULTIMODAUX.....	58
FIGURE 4.1 : SYSTEME AUTOMATIQUE DE RECONNAISSANCE DES EXPRESSIONS DE VISAGES EN 3D EN UTILISANT LA CAMERA KINECT (AREFK).	66

Liste des figures

FIGURE 4.2 : DETAILS DE L'ETAPE « ACQUISITION ».....	67
FIGURE 4.3 : DETAILS DE L'ETAPE « PRETRAITEMENT ».....	67
FIGURE 4.4 : DETAILS DE L'ETAPE « DETECTION ».....	68
FIGURE 4.5 : DETAILS DE L'ETAPE « CARACTERISATION ».....	69
FIGURE 4.6 : EXEMPLE DE LA METHODE MB-LBP.....	70
FIGURE 4.7 : MONTRE EXEMPLE DE LA METHODE MB_LBP POUR UN VISAGE.....	70
FIGURE 4.8 : DETAIL DE L'ETAPE « APPRENTISSAGE ».....	70
FIGURE 4.9 : DETAIL DE L'ETAPE « RECONNAISSANCE ».....	72
FIGURE 4.10 : LE CAPTEUR KINECT DE MICROSOFT.....	72
FIGURE 4.11 : MIRE DE CALIBRAGE D'OPENCV.....	74
FIGURE 4.12 : VISION HUMAINE STEREOSCOPIQUE.....	75
FIGURE 4.13 : RECONSTRUCTION D'UN MODELE DE VISAGE 3D.....	76
FIGURE 4.14 : DESSIN DE L'EXPRESSION PRESENTEE PAR L'UTILISATEUR VERS L'AVATAR.....	76
FIGURE 4.15 : LES IMAGES DE PROFONDEUR AVEC LE JEU DE DONNEES FRGCV1.0.....	77
FIGURE 4.16 : VUE DE NOTRE SYSTEME PROPOSE.....	78
FIGURE 4.17 : EXEMPLE DE L'IMAGE DE PROFONDEUR ET D'INTENSITE.....	80
FIGURE 4.18 : BALAYAGES DE VISAGE 3D ET LES POINTS SIFT DETECTES AVEC DES INFORMATIONS DE PROFONDEUR 3D.....	83
FIGURE 4.19 : MULTI-ECHELLES LBP DE L'IMAGE DE PROFONDEUR ET D'INTENSITE DU VISAGE.....	84
FIGURE 4.20 : POINTS-CLES DETECTES AVEC SIFT ET LES QUATRE IMAGES LBPP,R ASSOCIEES.....	84
FIGURE 4.21 : SCHEMA DE LA FUSION DES SCORES.....	85
FIGURE 5.1 : DETECTION DE VISAGE ET DESSIN DE RECTANGLE ENGLOBANT DANS CHAQUE TRAME.....	89
FIGURE 5.2 : SEGMENTATION SPATIALE DU VISAGE.....	90
FIGURE 5.3 : MEMBRES DE VISAGE LOCALISES ET ENREGISTRES DANS DES IMAGES.....	91
FIGURE 5.4 : PSEUDO-ALGORITHME DE LA PROCEDURE DE COLLECTE.....	92
FIGURE 5.5 : ALIGNEMENT DES DEUX FLUX CAMERA ..	92
FIGURE 5.6 : EXTRACTION DE VISAGE DES DEUX FLUX.....	93
FIGURE 5.7 : QUELQUES IMAGES D'EXPRESSIONS DE NOTRE BASE D'APPRENTISSAGES.....	94
FIGURE 5.8 : CERTAINS EIGENEXPRESSIONS CALCULEES A PARTIR DE NOTRE BASE.....	94
FIGURE 5.9 : CALCUL DE L'ORDONNEE Y D'UN POINT A PARTIR DE Z.....	95
FIGURE 5.10 : SEPT MODELES DE VISAGE 3D GENERIQUES.....	96
FIGURE 5.11 : INTERFACE DE VISUALISATION DE FLUX CAPTURE PAR LA CAMERA ..	96
FIGURE 5.12 : EXTRAIT DU REPERTOIRE « SAVEDEXPRESSIONS ».....	97
FIGURE 5.13 : FLUX DE VISUALISATION D' INTERFACE CAPTURE PAR KINECT.....	98
FIGURE 5.14 : INTERFACE PERMETTANT LA RECONNAISSANCE DU VISAGE AINSI SON EXPRESSION DETECTE.....	99
FIGURE 5.15 : FACIAL EXPRESSION RESEARCH GROUP BASE DE DONNEES (FERG-DB).....	100
FIGURE 5.16 : EXEMPLE DE LA BASE DE DONNEES FRGC v1.0.....	103
FIGURE 5.17: REPARTITION DES VALEURS ENTRE LES POINTS CLES DES DESCRIPTEURS SIFT DETECTES.....	104
FIGURE 5.18 : DISTRIBUTION DES VALEURS ENTRE LES POINTS-CLES DETECTES DANS DIFFERENTES SCANS	105
FIGURE 5.19 : TRACE LES COURBES CMC OBTENUES.....	106
FIGURE 5.20 : COURBE ROC.....	106
FIGURE 5.21 : COMPARAISON DES TEE POUR SIFT+LBP, SIFT ,LBP, ACP DANS L'ENSEMBLE D'EVALUATION.....	109

Liste des tableaux

TABLEAU 2.1 : DESCRIPTION DES SEPT EXPRESSIONS	43
TABLEAU 2.2 : DESCRIPTION DES SEPT EXPRESSIONS (SUITE)	44
TABLEAU 5.1 : TABLEAU DE NOTRE BASE DE DONNEES DE RECONNAISSANCE	101
TABLEAU 5.2 : PRECISION MOYENNE DE RECONNAISSANCE D'EXPRESSION(%) POUR CHAQUE EXPRESSION A TRAVERS TOUTE EXPRESSION EN UTILISANT UN EXPERT HUMAIN ET AREFK.	102
TABLEAU 5.3 : MONTRE LA PERFORMANCE DE NOTRE BASE DE DONNEES DE RECONNAISSANCE ET D'AUTRES BASES DE DONNEES PUBLIEES.....	102
TABLEAU 5.4 : PERFORMANCE DE LA METHODE D'EXTRACTION DE CARACTERISTIQUES LBP DANS L'ENSEMBLE D'EVALUATION ET DE TEST.....	108
TABLEAU 5.5 : PERFORMANCE DE METHODE D'EXTRACTION DE CARACTERISTIQUES LBP + SIFT DANS L'ENSEMBLE D'EVALUATION ET DE TEST	108
TABLEAU 5.6 : PERFORMANCES DES METHODES SIFT+LBP, SIFT, LBP, ACP.	110
TABLEAU 5.7 : MONTRE LA PERFORMANCE DE LA METHODE PROPOSEE ET D'AUTRES RESULTATS PUBLIES DE L'INDICE DE REFERENCE.....	110
TABLEAU 5.8 : CLASSIFICATION DES RESULTATS POUR LE JEU DE DONNEES FRGC V1.0 AVEC LA METHODE SIFT + LBP ET COMPARAISON AVEC NOS AUTRES METHODES UTILISEES.....	111

Introduction générale

1. Cadre de la thèse

De nos jours, on parle de plus en plus d'insécurité dans divers secteurs ainsi que des moyens informatiques à mettre en œuvre pour pallier les problèmes qui y sont associés : on peut citer le contrôle d'accès aux ordinateurs, le commerce électronique, les opérations bancaires basées sur l'identification du demandeur, la protection civile, les questions liées à la criminalité, le terrorisme, et tout récemment la lutte contre les fraudes sociales, etc. Afin de renforcer la sécurité dans tous ces domaines, devenue aujourd'hui une préoccupation internationale sous l'effet de la crise économique, la biométrie se présente comme une technologie potentiellement puissante. À l'origine, le mot « biométrie » désigne l'application au domaine de la biologie des mesures utilisées en mathématiques.

En effet, les différents moyens biométriques visent à utiliser des caractéristiques physiologiques et/ou comportementales spécifiques de chaque personne. De plus, nombre de ces modalités peuvent donner naissance à d'autres modalités lorsque le capteur est modifié. Par exemple, lors de la capture d'un visage à l'aide d'une caméra infrarouge ou une capture 3D, il s'agit toujours du visage, mais on parle alors de modalité infrarouge ou 3D du visage. Ces caractéristiques comportementales et/ou physiologiques présentent l'avantage d'être universelles, uniques et permanentes. Par ailleurs, elles ne peuvent pas être falsifiées à la différence des moyens classiques tels que les mots de passe ou les badges.

Un système biométrique est essentiellement un système de reconnaissance de formes qui utilise les données biométriques d'un individu. Selon le contexte de l'application, un système biométrique peut fonctionner en mode d'enrôlement ou en mode de vérification ou bien en mode d'identification.

Durant les vingt dernières années, la reconnaissance automatique de visages est devenue un enjeu primordial, notamment dans les domaines de la sécurité et de l'indexation de documents multimédia. Malgré la multitude de techniques proposées et les progrès réalisés, le problème reste non entièrement résolu, comme le montre l'évaluation menée par le National Institute of Standards and Technology (NIST).

2. Problématique

Plusieurs méthodes ont été développées pour la reconnaissance bidimensionnelle des visages. Elles véhiculent, cependant, un certain nombre de difficultés qui proviennent essentiellement de la forte variabilité des données à classer dues à des changements dans l'expression faciale, dans la pose de la tête, dans les conditions d'illumination, à des occultations partielles et dans le fait que les images de visages constituent des données de très grandes dimensions. Ces dernières années, on parle de plus en plus des techniques de reconnaissance de visage 3D comme une solution alternative pour résoudre les problèmes cités ci-dessus. En effet, la richesse de l'information fournie par les mesures 3D permet de reconstruire la forme

tridimensionnelle du visage. Ce type de représentation du visage est invariant aux changements d'illumination et de pose.

Dans notre travail, nous nous sommes proposés de développer une technique de reconnaissance bidimensionnelle de visages pour proposer ensuite une technique de reconnaissance tridimensionnelle de visages. Notre thèse est constituée d'une introduction générale suivie de cinq chapitres et d'une conclusion générale.

3. Organisation de la thèse et Contributions

Dans le premier chapitre, nous rappelons des concepts de base en biométrie faciale. Nous tentons également de justifier le choix de la reconnaissance du visage en particulier. Les faiblesses des systèmes de reconnaissance automatique de visage 2D. Nous introduisons ensuite la reconnaissance de visage 3D à travers la présentation des techniques 3D et des bases de données 3D de visage disponibles.

Le deuxième chapitre présente l'état de l'art sur les deux notions complémentaires de nos travaux, d'une part la reconnaissance automatique de visages 2D et, d'autre part, la reconnaissance 3D de visages. Tout d'abord, on commence par un bilan des importants algorithmes de reconnaissance de visage 2D, puisqu'en général les algorithmes de reconnaissance de visage 3D sont des extensions directes d'algorithmes 2D existants. Ce chapitre permet d'évaluer les capacités et les lacunes des algorithmes de reconnaissance de visage en cours afin d'identifier les problèmes et les orientations pour la recherche future dans ce domaine.

Dans le troisième chapitre, nous présentons notre approche de vérification unimodale et multi-algorithmes de visage 2D. Dans la partie de vérification de visage 2D, nous présentons tout d'abord notre modèle de vérification de visage. Nous détaillons ensuite les quatre principales étapes de ce modèle. Enfin, nous commençons l'étude expérimentale sur le système automatique 2D par la comparaison des algorithmes de réduction de l'espace de données, puis nous étudions l'extraction des caractéristiques par SIFT (Scale Invariant Feature Transform) plus MB_LBP (Multi Block_Local Binary Patterns). Dans la partie de vérification multi-algorithmes de visage 2D, nous passons en revue les limitations des systèmes biométriques lorsqu'ils n'utilisent qu'une seule modalité biométrique. Enfin, nous détaillons les résultats expérimentaux.

Dans le quatrième chapitre, nous testons la façon dont les caractéristiques locales de l'information bidimensionnelles et tridimensionnelles contribuent à la reconnaissance faciale lorsque l'éclairage, les changements d'expressions et la combinaison d'expression sous l'éclairage sont pris en compte. Nous présentons nos deux contributions pour la vérification de visage 3D, la première contribution contient l'étude théorique de notre conception pour un système de reconnaissance de visage 3D et nous présentons quelques approches utilisées dans le système de reconnaissance faciale 3D.

Dans un premier temps, nous étudions successivement, les objectifs de notre système tout en donnant une description détaillée sur le système et nous présenterons les différentes étapes pour atteindre l'objectif fixé dans le cadre de notre projet. Finalement, nous décrivons la manière d'évaluation de notre système.

Introduction générale

Pour la deuxième contribution, nous présentons dans un premier temps la phase de prétraitement par l'alignement (translation et rotation) de l'image tridimensionnelle en entrée par rapport à l'image de référence (image du visage avec une expression neutre) en utilisant l'algorithme PCA (Principal Component Analysis). Selon les images alignées, nous obtenons les images tridimensionnelles de profondeur et des images d'intensité 2D. Ensuite, nous détaillons les méthodes d'extraction des caractéristiques utilisées et notre descripteur SIFT+LBP, puis la méthode de réduction d'espace basée sur les SVM (Support Vector Machine). Enfin, nous testons notre approche sur les bases de données FRGC V1.0.

Le cinquième chapitre est consacré à l'implémentation des différentes étapes décrites dans les précédents chapitres. Dans un premier temps, nous étudions successivement, les objectifs de notre système tout en donnant une description détaillée sur le système et nous présentons les différentes étapes pour atteindre l'objectif fixé dans le cadre de notre projet. Enfin, nous décrivons la manière d'évaluation de notre système. Nous terminons ce mémoire par une conclusion générale et les perspectives futures potentielles.

Introduction à la biométrie

1.1 Introduction

Dans ce chapitre, nous mettons en relief quelques notions et définitions de base liées à la biométrie et ses diverses technologies. Nous analysons le principe de fonctionnement des systèmes biométriques, les outils utilisés pour mesurer leurs performances ainsi que les domaines d'applications. Nous insistons ici surtout sur la place qu'occupe la reconnaissance faciale parmi les autres techniques biométriques, du fait qu'elle constitue l'objectif central de notre travail. Enfin, nous relatons les difficultés majeures liées à la reconnaissance bidimensionnelle de visages 2D tout en passant en revue les bases de données de visage 2D disponible. Nous introduisons, par ailleurs, la reconnaissance bidimensionnelle de visages à travers la présentation des techniques 3D ainsi que les bases de données de visage 3D disponibles.

1.1.1 Définition de la biométrie

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

1.1.2 Panorama des différentes techniques biométries [1]

Les systèmes biométriques sont généralement classés en trois grandes catégories : la biométrie morphologique ou physiologique, la biométrie comportementale et la biométrie biologique ou émergente.

La biométrie morphologique est basée sur l'identification de traits physiques particuliers qui, pour toute personne, sont uniques et permanents. Cette catégorie regroupe la reconnaissance des empreintes digitales, la forme de la main, la forme du visage, la rétine et l'iris de l'œil et les veines de la main [2].

La biométrie comportementale se base sur l'analyse de certains comportements d'une personne comme le tracé de sa signature, l'empreinte de sa voix, son mouvement de lèvre ou encore sa façon de taper sur un clavier [2,3].

L'ADN, la forme des oreilles, l'odeur du corps humain et l'analyse de la démarche sont également étudiés, mais ses caractéristiques sont considérées comme émergentes (stade purement expérimental) [3].

En théorie, on dit qu'un critère physiologique ou comportemental est exploitable en reconnaissance d'individus s'il satisfait aux conditions ci-après [4] :

- Universalité : toutes les personnes d'une population à identifier doivent la posséder ;

Chapitre 1: Introduction à la biométrie

- Unicité : deux personnes ne doivent pas posséder exactement la même caractéristique ;
- Permanence : cela signifie qu'elle ne doit pas varier au cours du temps ;
- Mesurabilité : qui justifie de la possibilité de le quantifier.

En pratique, il existe d'autres conditions importantes :

- Performance : l'identification doit être précise et rapide ;
- Acceptation : elle doit être bien acceptée par les utilisateurs du système ;
- Circonvention : qui réfère à la facilité ou la difficulté de le pirater [5].



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

1.1.2.1 Pourquoi la reconnaissance de visage ?

C'est la caractéristique biométrique qu'utilisent le plus les humains pour s'identifier entre eux, ce qui explique qu'elle est généralement acceptée par les utilisateurs. Le système d'acquisition est soit un appareil photo ou une caméra numérique. La difficulté de reconnaissance de visage varie énormément suivant que l'acquisition se fasse dans un environnement contrôlé ou non. Dans un environnement contrôlé, des paramètres tels que l'arrière-plan, la direction et l'intensité des sources lumineuses, l'angle de la prise de vue, la distance de la caméra au sujet sont des paramètres maîtrisés par le système. Dans un environnement non contrôlé, une série de prétraitements sont souvent nécessaires avant de pouvoir faire la reconnaissance proprement dite. Il faut tout d'abord détecter la présence ou l'absence de visage dans l'image, le visage devant ensuite être segmenté. Enfin, si nous travaillons sur un flux vidéo, le système doit suivre le visage d'une image à l'autre [1].

La performance des systèmes de reconnaissance du visage a connu une nette amélioration, mais les résultats sont encore loin d'être atteints. Les évaluations organisées par le NIST1 (National Institute of Standards and Technology) sur la base de données FERET ont montré que les systèmes sont très sensibles aux variations d'illumination et de pose [6]. De plus, si sur une période de temps relativement court, les systèmes de reconnaissance du visage ont en général des performances acceptables [7].

Un ensemble de scans du visage 3D de sujets connus sont initialement stockés. Cet ensemble est généralement appelé ensemble de galeries (gallery). Lors de la reconnaissance, les scans du visage 3D des sujets appartiennent à l'ensemble de la galerie où d'autres sujets sont appelés sondes (probes), utilisées pour correspondre avec les scans dans l'ensemble de galeries. Dans notre travail, nous étudions la vérification de visage 3D, qui constitue par ailleurs l'application la plus difficile.

La forme et la texture sont deux propriétés intrinsèques d'un objet du visage. Heureusement, le système actuel d'acquisition de visage 3D peut généralement offrir des scans du visage 3D (forme) avec leurs textures correspondantes. Dans ce cas, une image 2D est aussi appelée « image 2D de texture » [4].

Un visage numérisé 3D muni de sa carte de texture est généralement appelé « visage 3D texturé ». Une analyse de visage 3D peut être rendue sous forme d'image de profondeur, de nuages de points ou de maillage triangulaire. L'image de profondeur est une image dans laquelle la valeur du pixel enregistre la distance entre le capteur et la surface du visage. Les nuages de points enregistrent la forme de la géométrie d'un modèle de visage 3D en coordonnées 3D. Le maillage triangulaire convertit la forme de la géométrie d'un modèle de visage 3D en points 3D et leurs structures de topologie. La figure 1.2 illustre un exemple de l'ensemble de ces quatre types de formats d'enregistrement de visage [4].

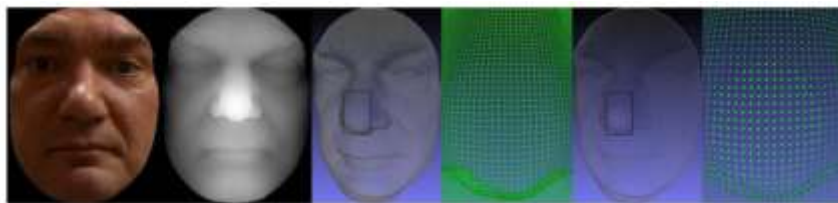


Figure 1.2. Quatre types de formats d'enregistrement du visage 3D, de gauche à droite : texture de visage 3D, image de profondeur, maille en forme de triangle et la zone de nez zoomé, nuages de points et la région de nez zoomée [4].

1.1.2.2 *Qu'est-ce que la Multimodalité ?*

Bien que de nos jours, il existe des techniques biométriques extrêmement fiables telles que la reconnaissance de la rétine ou de l'iris, elles sont cependant coûteuses, intolérables par le public et ne peuvent être réservées qu'à des applications de haute sécurité. Néanmoins, les techniques telles que la reconnaissance du visage ou la voix sont acceptables par les utilisateurs ; simplement, elles ont des performances peu satisfaisantes pour être diffusées dans des conditions réelles. Afin d'améliorer la sécurité des systèmes énumérés précédemment, on doit au préalable résoudre l'intégration de la biométrie avec l'identification basée sur une connaissance ou une possession. Cette méthode permet d'améliorer la sécurité du système, mais possède des faiblesses inhérentes à l'identification basée sur une connaissance ou une possession [4].

La multimodalité est une alternative permettant l'amélioration, de manière systématique, de la performance d'un système biométrique. On entend par performance, la précision du système et son efficacité, plus particulièrement en matière d'identification. En effet, les différents classificateurs

font généralement des erreurs, ce qui nous permet de tirer profit de cette complémentarité afin d'améliorer la performance globale du système. Les différentes formes de multimodalités sont les suivantes :

1. Systèmes multiples biométriques : exemple on combine la reconnaissance du visage, celle des empreintes digitales et la reconnaissance du locuteur. C'est le sens le plus classique du terme multimodal.
2. Systèmes multiples d'acquisition : on cite comme exemple l'utilisation de deux scanners différents (l'un optique, l'autre thermique) pour la reconnaissance d'empreintes digitales.
3. Mesures multiples d'une même unité biométrique : par exemple, faire la reconnaissance des deux iris ou des dix doigts d'un même individu.
4. Instances multiples d'une même mesure : faire une capture répétée du même attribut biométrique avec le même système d'acquisition.
5. Algorithmes multiples : utiliser différents algorithmes de reconnaissance sur le même signal d'entrée.

Il convient de comparer les mérites respectifs de chacune des formes de multimodalité. Les critères retenus se distinguent par la différence du coût, la gêne pour l'utilisateur et la quantité d'information par rapport au système biométrique unimodal.

Selon les modalités de visage adopté pour la galerie et la sonde, la reconnaissance faciale 3D peut être effectuée dans trois scénarios correspondants : reconnaissance de visage uniquement 3D de forme à base (forme 3D contre forme 3D), reconnaissance faciale multimodale 2D+3D (3D texturée contre 3D texturée) et la reconnaissance de visage 3D asymétrique (forme 3D contre texture 2D ou vice versa). La reconnaissance de visage multimodale (2D+3D) s'est avérée être plus précise que la reconnaissance uniquement 3D de forme à base de visage [25]. Dans ce travail, nous nous concentrons sur le scénario de reconnaissance faciale multimodale 2D+3D. [4]

1.1.3 Les systèmes biométriques [1]

C'est effectivement un système de reconnaissance des formes. Il utilise les données biométriques d'un individu. Il existe au moins deux modules dans un système biométrique : le module d'apprentissage et celui de reconnaissance. Le troisième module (facultatif) est le module d'adaptation [7,8] :

- ❖ Le module d'enrôlement : est une phase d'apprentissage qui a pour but de recueillir des informations biométriques sur les personnes à identifier. Durant cette phase, les caractéristiques biométriques des individus sont saisies par un capteur biométrique, puis représentées sous forme numérique (signatures), et enfin stockées dans la base de données. Il est à noter que la qualité du capteur peut grandement influencer les performances du système. Meilleure est la qualité du système d'acquisition, moins il y aura des prétraitements à effectuer pour extraire les paramètres du signal. Cependant, les capteurs de qualité sont généralement coûteux, leur utilisation est donc limitée à des applications de haute sécurité

pour un public restreint. Le traitement lié à l'enrôlement n'a pas de contraintes de temps, du fait qu'il s'effectue « hors ligne ».

- ❖ Le module de reconnaissance : Durant cette étape, la caractéristique biométrique est mesurée, un ensemble de paramètres est extrait comme lors de l'étape d'apprentissage. Le capteur utilisé doit disposer de propriétés aussi proches que possible du capteur utilisé durant la phase d'apprentissage. Si les deux capteurs ont des propriétés trop différentes, nous serons obligés de procéder à une série de prétraitements supplémentaires pour limiter la dégradation des performances. La suite de la reconnaissance sera différente suivant le mode opératoire du système : identification ou vérification.
- Le mode de vérification ou authentification : est un mode de comparaison "un à un", dans lequel le système valide l'identité d'une personne en comparant les données biométriques saisies avec le modèle biométrique de cette personne stockée dans la base de données du système. Dans un tel mode, le système doit alors répondre à la question suivante : « Suis-je réellement la personne que je suis en train de proclamer ? ».
- Le mode d'identification : est une comparaison "1 à N", dans lequel le système reconnaît un individu en l'appariant avec un des modèles de la base de données. La personne peut ne pas être dans la base de données. Ce mode consiste à associer une identité à une personne. En d'autres termes, il répond à des questions du type : « Qui suis-je ? ».

1.1.3.1. Mesure de la performance d'un système biométrique

Tout d'abord, afin de comprendre comment déterminer la performance d'un système biométrique, il nous faut définir clairement trois critères principaux :

- 1) Taux de Faux Rejet (TFR) ou (False Reject Rate, FRR) : ce taux représente le pourcentage de personnes censées être reconnues, mais qui sont rejetées par le système.

$$FRR = \frac{\text{Nombre de faux rejets}}{\text{Nombre total d'accès client}} \quad (1.1)$$

- 2) Taux de Fausse Acceptation (TFA) ou (False Accept Rate, FAR) : ce taux représente le pourcentage de personnes censées ne pas être reconnues, mais qui sont tout de même acceptées par le système.

$$FAR = \frac{\text{Nombre de fausses acceptations}}{\text{Nombre total d'accès client}} \quad (1.2)$$

- 3) Taux d'Égale Erreur TEE ou (Equal Error Rate, EER) : Un autre critère de performance (taux de reconnaissance) est souvent déterminé, il équivaut à $(100 - FAR - FRR)$, exprimée en pourcentage. Ce taux est calculé à partir des deux premiers critères et constitue un point de mesure de performance courant. Ce point correspond à l'endroit où $FRR = FAR$, c'est-à-dire le meilleur compromis entre les faux rejets et les fausses acceptations.

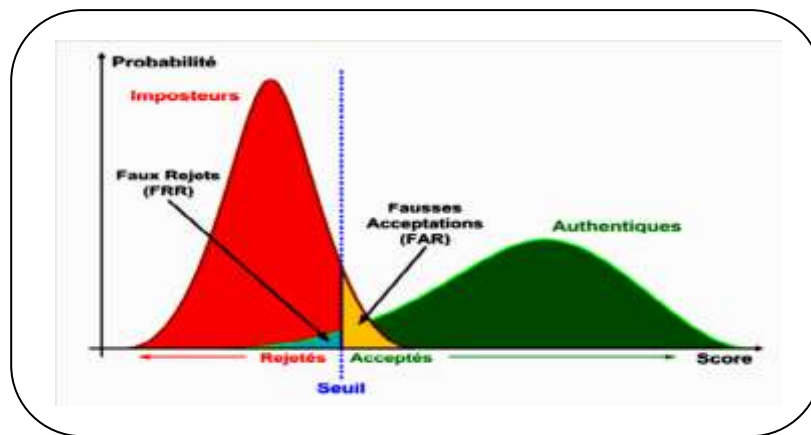


Figure 1.3. Illustration du FRR et du FAR.

Selon la nature (authentification ou identification) du système biométrique, il existe deux façons d'en mesurer la performance :

– Lorsque le système opère en mode authentification, on utilise ce que l'on appelle une courbe **ROC** (pour "**R**eceiver **O**perating **C**haracteristic" en anglais). La courbe ROC (Figure. I.4) trace le taux de faux rejet en fonction du taux de fausse acceptation. Plus cette courbe tend à épouser la forme du repère, plus le système est performant, c'est-à-dire possédant un taux de reconnaissance global élevé.

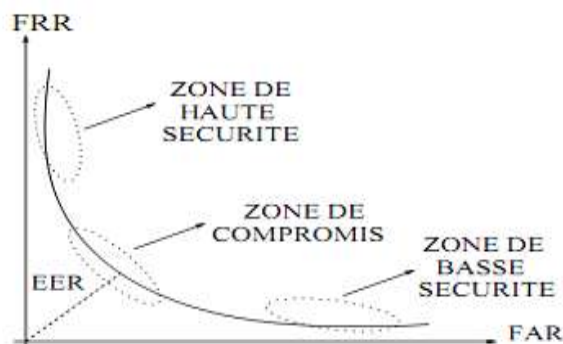


Figure 1.4. Courbe ROC [9].

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

- 1) Taux de reconnaissance (Rank-one Recognition Rate) : Il mesure le pourcentage des entrées qui sont correctement identifiées.
- 2) Cumulative Match Characteristic (CMC) : En revanche, dans le cas d'un système utilisé en mode identification, on utilise ce que l'on appelle une **courbe CMC** (pour "**C**umulative **M**atch **C**haracteristic" en anglais). La courbe CMC (Figure I.5) donne le pourcentage de personnes reconnues en fonction d'une variable que l'on appelle le rang [Beve 05]. On dit qu'un système reconnaît au **rang 1** lorsqu'il choisit la plus proche image comme résultat de la reconnaissance. On dit qu'un système reconnaît au **rang 2**, lorsqu'il choisit, parmi deux images, celle qui correspond le mieux à l'image d'entrée, etc. On peut donc dire que plus le rang s'augmente, plus le taux de reconnaissance correspondant est lié à un niveau de sécurité faible.

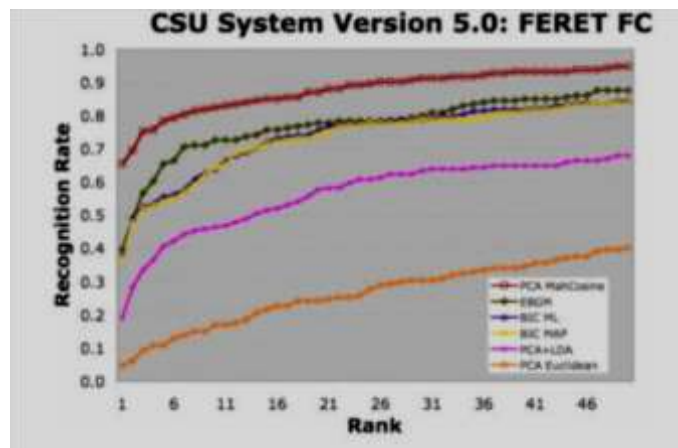


Figure 1.5. Courbes CMC du CSU System 5.0 pour le “FERET Probe Set FC” et pour différents algorithmes de reconnaissance faciale.

Le taux de reconnaissance et la courbe CMC (Cumulative Match Characteristic) sont couramment utilisés pour évaluer l'identification de visage 3D. Le TEE (Taux d'Égale Erreur) et la courbe ROC sont couramment utilisés pour évaluer la vérification de visage 3D.

1.1.4 Les modules d'un système biométrique

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

- ❖ Le module d'adaptation : Durant la phase d'apprentissage, le système biométrique ne capture souvent que quelques instances d'un même attribut afin de limiter la gêne pour l'utilisateur. Il est donc difficile de construire un modèle assez général capable de décrire toutes les variations possibles de cet attribut. De plus, les caractéristiques de cette biométrie ainsi que ses conditions d'acquisition peuvent varier. L'adaptation est donc nécessaire pour maintenir voire améliorer la performance d'un système utilisation après utilisation. L'adaptation peut se faire en mode supervisé ou non supervisée, mais le second mode est de loin le plus utile en pratique. Si un utilisateur est identifié par le module de reconnaissance, les paramètres extraits du signal serviront alors à réestimer son modèle.

En général, le taux d'adaptation dépend du degré de confiance du module de reconnaissance dans l'identité de l'utilisateur. Bien entendu, l'adaptation non supervisée peut poser un problème en cas d'erreurs du module de reconnaissance. L'adaptation est quasi indispensable pour les caractéristiques non permanentes comme la voix.

Les différents modules qui composent un système biométrique sont représentés sur la Figure 1.6; leur fonctionnement est assuré par les modules suivants :

- ❖ Module capteur biométrique : correspond à la lecture de certaines caractéristiques physiologiques, comportementales ou biologiques d'une personne, au moyen d'un terminal de capture biométrique (ou capteur biométrique) ;
- ❖ Module extraction des données : extrait les informations pertinentes à partir des données biométriques brutes, par exemple des images du visage ou des régions caractéristiques du visage ;
- ❖ Module création d'une signature : crée un modèle numérique afin de représenter la donnée biométrique acquise. Ce modèle, appelé aussi signature, sera conservé sur un support portable (puce ou autre) ou dans une base de données ;
- ❖ Module comparaison : compare les caractéristiques biométriques d'une personne soumise au contrôle (volontairement ou à son insu) avec les « signatures » mémorisées. Ce module fonctionne soit en mode vérification (pour une identité proclamée) ou en mode identification (pour une identité recherchée).
- ❖ Module base de données : stocke les modèles biométriques des utilisateurs enrôlés.

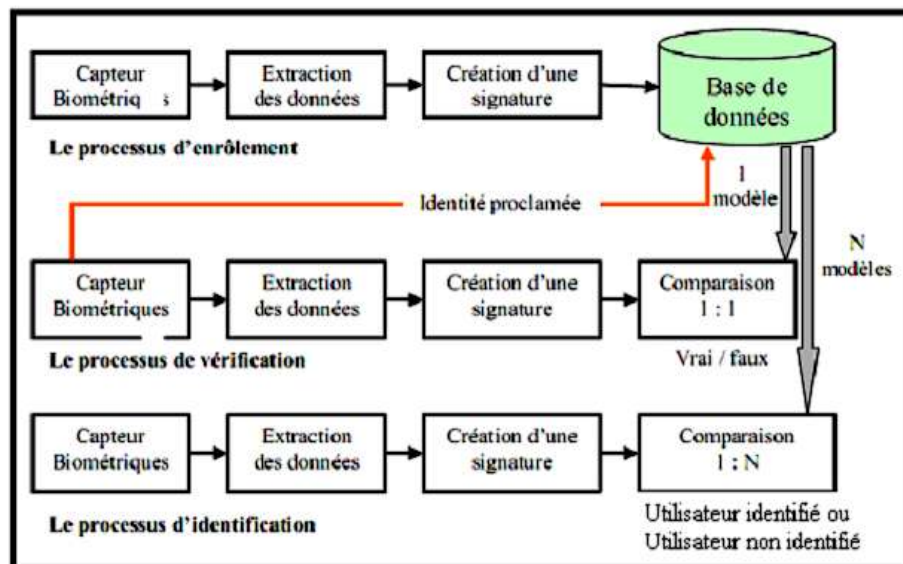


Figure 1.6. Principaux modules d'un système biométrique ainsi que les différentes modes [8].

1.1.5 Rôle de la reconnaissance de visage

Les visages constituent une catégorie de stimulus unique par la richesse des informations qu'ils véhiculent. Ils sont à la fois les vecteurs visuels principaux de l'identité individuelle, et des vecteurs essentiels de communication (verbale et non verbale) d'intentions et d'émotions entre individus, via, en particulier, la direction du regard et les expressions faciales. Sur le plan perceptif, la reconnaissance des visages constitue une des capacités les plus étonnantes, et les plus performantes du système visuel humain. En effet, l'être humain est capable de discriminer et de reconnaître un très grand nombre de visages, alors même que les visages constituent une catégorie de stimulus homogène, partageant un ensemble de traits similaires et une structure commune, et qu'ils peuvent subir d'importants changements avec l'âge par exemple, ou encore selon l'expression faciale [10].

Le but de la reconnaissance des visages est de concevoir des systèmes informatiques capables de copier les prodigieuses facultés de reconnaissance du cerveau humain. Cependant, et malgré

l'avancement de la recherche réalisé durant ces trois dernières décennies, la reconnaissance robuste de visages reste très difficile comme le montre la récente évaluation menée par le National Institut of Standards and Technology (NIST) [11].

Les difficultés proviennent essentiellement de la forte variabilité des données à classer (dues à des changements dans les expressions faciales, dans la pose de la tête, dans les conditions d'illumination et à des occultations partielles), et dans le fait que les images de visages constituent des données de très grandes dimensions. Néanmoins, et malgré toutes ces difficultés, le domaine de la reconnaissance faciale demeure important pour de nombreux chercheurs vu la multitude des champs d'applications. Cette variabilité des applications a donné lieu à la disponibilité de plusieurs produits commerciaux permettant :

- L'identification, par comparaison d'un visage à ceux mémorisés dans une base de données.
- La vérification, par comparaison des identités déclarées avec les identités associées aux visages mémorisés.
- La supervision, qui permet de suivre l'image d'une personne dans une séquence vidéo [12].
- La surveillance qui permet de retrouver, en temps réel, une personne dans une séquence vidéo à partir d'une liste de visages.

À l'inverse, des systèmes d'identification basés sur l'analyse des images du visage ne présentent aucune contrainte pour les utilisateurs. La reconnaissance faciale est une modalité qui peut être implémentée de manière indépendante des autres modalités biométriques, elle est souvent utilisée dans des applications de surveillance. La reconnaissance faciale offre plusieurs avantages :

Le système de capture (c.-à-d. les caméras) est facile à installer, il est accepté dans les lieux publics ce qui permet d'avoir des bases de données de plus en plus grandes et ainsi améliorer les performances de la reconnaissance. Ces derniers temps, la reconnaissance faciale suscite un intérêt croissant auprès de la communauté scientifique, qui s'est notamment manifesté à travers l'organisation de conférences internationales spécialisées telle que « The International Conference on Audio and Video-based Biometric Person Authentication (AVBPA) » depuis 1997, et « the International Conference on Automatic Face and Gesture Recognition (AFGR) », depuis 1995. Par ailleurs, les États-Unis ont mené depuis 1993 une série de tests de reconnaissance faciale dont les résultats sont accessibles au public. Ces tests sont désignés sous les noms de FERET, XM2VTS, FRVT 2000 et FRVT 2002 [8].

1.1.6 Systèmes biométriques basés sur la reconnaissance de visage

Tout processus automatique de reconnaissance des visages doit prendre en considération plusieurs facteurs qui contribuent à la complexité de sa tâche, car le visage est une entité dynamique qui change constamment sous l'influence de plusieurs facteurs. La Figure 1.7 illustre la démarche généralement adoptée pour réaliser de tels systèmes [13].

- Le monde physique : trois paramètres sont considérés : L'éclairage, la variation de posture et l'échelle. La variation de l'un de ces trois paramètres peut conduire à une distance entre deux images du même individu, supérieure à celle séparant deux images de deux individus différents.
- Le codage : consiste en l'acquisition de l'image et sa digitalisation, il comporte un risque de bruit et donne lieu à une représentation 2D (la matrice des niveaux de gris) pour un objet 3D (le visage). Peu de méthodes 3D ont été proposées, mais cela devrait changer dans un avenir

très proche avec l'invention des Range-Cameras qui peuvent mesurer la topologie 3-D du visage.

- Le prétraitement : On doit éliminer le bruit par des techniques de traitement et de restauration d'images et procéder à une localisation de la tête dans l'image, cette opération est très complexe, beaucoup supposent que l'image contient un visage et que l'arrière-plan est neutre. La restauration d'images ou l'élimination du bruit consiste à compenser les dégradations connues ou estimées et rétablir la qualité initiale de l'image [13].

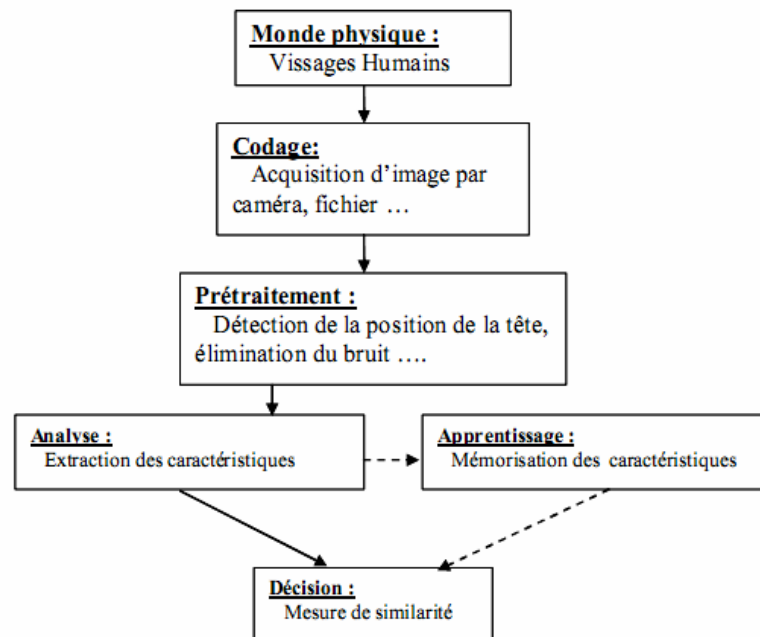


Figure 1.7. Le schéma général d'un système de reconnaissance de visage [13].

L'efficacité des systèmes biométriques basés sur l'authentification de visage dépend essentiellement de la méthode utilisée pour localiser le visage dans l'image. Dans la littérature scientifique, le problème de localisation de visages est aussi désigné par la terminologie "détection de visages". Plusieurs travaux de recherches ont été effectués dans ce domaine. Ils ont donné lieu au développement d'une multitude de techniques allant de la simple détection du visage, à la localisation précise des régions caractéristiques du visage, tels que les yeux, le nez, les narines, les sourcils, la bouche, les lèvres, les oreilles, etc.

Cependant, les solutions proposées jusqu'à présent sont loin d'être satisfaisantes, car elles fonctionnent uniquement dans des environnements contrôlés, et par conséquent elles ne gèrent pas la variabilité des conditions d'acquisition de la vie quotidienne, notamment [8] : La pose, la présence ou l'absence des composantes structurelles, les occultations, les conditions d'illumination.

- L'étape d'analyse : appelée aussi indexation, représentation, modélisation ou extraction de caractéristiques, on doit extraire de l'image les informations qui seront sauvegardées en mémoire pour être utilisées plus tard dans la phase Décision. Le choix de ces informations utiles revient à établir un modèle pour le visage, elles doivent être discriminantes et non redondantes [13].

- L'apprentissage : consiste à mémoriser les représentations calculées dans la phase d'analyse pour les individus connus.
- La décision : Pour estimer la différence entre deux images, il faut introduire une mesure de similarité [13].

1.1.7. Mode de reconnaissance

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

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

1.2 La reconnaissance de visage

1.2.1 Processus de reconnaissance des visages

1.2.1.1 Images statiques

La plupart des algorithmes de reconnaissance automatique de visages portent sur la classification d'images statiques 2D notamment en niveau de gris. Les algorithmes prenant en compte la couleur ne sont, pour la plupart, que des généralisations à trois canaux des techniques utilisées en niveaux de gris. On dispose d'une base de connaissances contenant un ensemble d'images de visages de personnes connues, c'est-à-dire les personnes que le système est censé reconnaître. Cette base de connaissances peut contenir plusieurs images d'une même personne sous des conditions de prises de vue différentes. Chaque image est étiquetée par son identité associée. Lorsqu'un visage requête (image d'un visage à reconnaître) est présenté au système, celui-ci va chercher à lui assigner une identité parmi celles contenues dans la base de connaissances. Le système de reconnaissance est basé sur un modèle généralement construit à partir d'une base d'apprentissage contenant un ensemble d'images de visages par le biais d'un algorithme d'apprentissage.

Dans certains cas, la base d'apprentissage et la base de connaissances sont confondues. Le modèle est alors spécifiquement conçu pour les visages qu'il vise à reconnaître. Dans d'autres cas, au contraire, les deux bases sont distinctes soit à cause de l'application à réaliser soit à cause des spécificités de la technique utilisée [14].

1.2.1.2. Images 3D

Très récemment, les avancées dans le domaine de l'acquisition de données tridimensionnelles (acquisition par scanners laser) ont permis l'émergence de technologies de reconnaissance 3D. Deux grandes familles de méthodes sont apparues ces dernières années :

1. Celles basées sur la correspondance d'objets 3D directement (on dispose d'images 3D aussi bien dans la base d'apprentissage que pour les visages requêtes) [15].
2. Celles se ramenant pour la classification à un problème de reconnaissance 2D (seules les vues 3D de la base d'apprentissage sont requises) [16].

L'inconvénient de ces deux méthodes c'est qu'elles sont intrusives en utilisant un capteur 3D.

1.2.2 Principales difficultés de la reconnaissance de visage

Pour le cerveau humain, le processus de la reconnaissance de visages est une tâche visuelle de haut niveau. Bien que les êtres humains puissent détecter et identifier des visages dans une scène sans beaucoup de peine, construire un système automatique qui accomplit de telles tâches représente un sérieux défi. Ce défi est d'autant plus grand lorsque les conditions d'acquisition des images sont très variables. Il existe deux types de variations associées aux images de visages : inter et intra sujet. La variation inter-sujet est limitée à cause de la ressemblance physique entre les individus. Par contre, la variation intra-sujet est plus vaste.

Elle peut être attribuée à plusieurs facteurs que nous analysons ci-dessous. Nous détaillons ici les principales difficultés rencontrées par un système de reconnaissance faciale automatique 2D dans des conditions réelles :

- ❖ **Illumination** : Les variations d'illumination viennent entraîner des variations considérables dans l'apparence d'un visage. Deux types d'éclairage peuvent influencer celle-ci : l'illumination globale (ou ambiante) et l'illumination locale. Alors que l'illumination globale affecte tout le visage de manière uniforme (ou presque), l'illumination locale entraîne la création de zones d'ombres et de zones éclairées, et ce, de manière non linéaire. La figure 1.8 présente un exemple de visage dont la source lumineuse l'éclairant se déplace. De nombreuses approches ont été proposées pour gérer ces problèmes de luminosité. Une modélisation implicite de la luminosité dans la création d'un modèle de visage peut être réalisée [1].

L'extraction de caractéristiques invariantes aux changements de luminosité est également une approche largement décrite dans la littérature. Notons enfin qu'un certain nombre d'approches traitent le problème de la luminosité en amont de la reconnaissance par une étape de prétraitement dont l'objectif principal est bien souvent de corriger les artefacts dus aux variations de luminosité.



Figure1.8. Exemple de variation d'éclairage [8].

- ❖ Pose : Le taux de reconnaissance de visage baisse considérablement quand des variations de pose sont présentes dans les images. Cette difficulté a été démontrée par des tests d'évaluation élaborés sur les bases FERET et FRVT [17,133].

La variation de pose est considérée comme un problème majeur pour les systèmes de reconnaissance faciale. Quand le visage est de profil dans le plan image (orientation $< 30^\circ$), il peut être normalisé en détectant au moins deux traits faciaux (passant par les yeux). Cependant, lorsque la rotation est supérieure à 30° , la normalisation géométrique n'est plus possible (voir Figure 1.9).



Figure 1.9. Exemples de variation de poses (Images recueillies à partir d'Internet).

- ❖ Expressions faciales : L'apparence d'un visage varie grandement en présence d'expressions faciales (Figure 1.10). Les éléments faciaux tels que la bouche ou encore les yeux peuvent alors subir des déformations importantes, pouvant faire échouer un système de reconnaissance faciale fondée par exemple sur des points d'intérêt (ceux-ci pouvant ainsi subir d'importantes translations). La bouche est en général l'élément facial qui varie le plus, mais l'aspect des sourcils peut par exemple être grandement modifié [9].

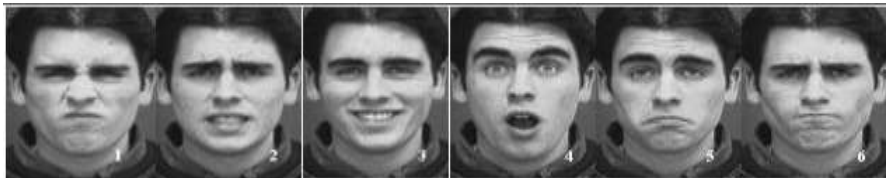


Figure 1.10. Exemples de variation d'expressions [8].

- ❖ Oclusions : Les oclusions partielles apparaissent fréquemment dans des applications réelles, comme illustré sur la figure 1.11. Elles peuvent être causées par une main cachant une partie du visage, par des cheveux longs, des lunettes de vue, de soleil, par tout autre objet (foulard . . .), ou encore par une autre personne. Il arrive également qu'une partie du visage en cache une autre, comme dans le cas d'une rotation hors plan par exemple.



Figure 1.11: Exemples d'occlusion du visage (Image recueillie à partir d'Internet).

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

Dans le contexte de la biométrie, les systèmes proposés doivent être non intrusifs, c'est-à-dire qu'on ne doit pas compter sur une coopération active du sujet. Par conséquent, il est important de savoir reconnaître des visages partiellement occultés.

1.2.3 Opportunités de reconnaissance de visage 3D [4]

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

- 1) L'invention et la vulgarisation des techniques de numérisation 3D : avec l'invention et le développement rapide des technologies de numérisation 3D, le monde connaît une évolution numérique 3D. Le mesurage, la reconstruction, la numérisation et la reproduction d'objets en 3D du monde réel sont de plus en plus faciles et populaires. De nos jours, de nombreux types de capteurs 3D sont disponibles pour capturer des données 3D statiques et même dynamiques de surface du visage. On peut se référer à la section 1.2.4.1 « Techniques d'acquisition de visage 3D » pour des exemples détaillés.
- 2) La disponibilité d'un grand potentiel pour lutter contre les difficultés 2D : les informations sur la forme des approches fondées sur la reconnaissance faciale en 3D sont naturellement moins sensibles aux conditions d'éclairage par comparaison avec les informations de photométrie 2D. La surface faciale 3D est plus commode pour effectuer la correction de pose et résoudre le problème de

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

3) La forme du visage 3D et la photométrie 2D contiennent des informations comparables et complémentaires : Comme indiqué dans [18], de nombreux résultats de l'étude fournissent la preuve que les indices photométriques 2D et de forme 3D sont utilisés à peu près par l'homme pour reconnaître les visages. En outre, nous sommes particulièrement intéressés à la reconnaissance faciale 3D, car il est communément admis que l'utilisation de la détection 3D a le potentiel pour une plus grande précision de la reconnaissance qu'en 2D.

D'autre part, il a été prouvé à maintes reprises que les performances de la reconnaissance de visage multimodale (2D +3D) sont toujours meilleures que celles en 2D seul. C'est-à-dire, la forme du visage et la photométrie sont aussi complémentaires les unes aux autres pour la reconnaissance faciale.

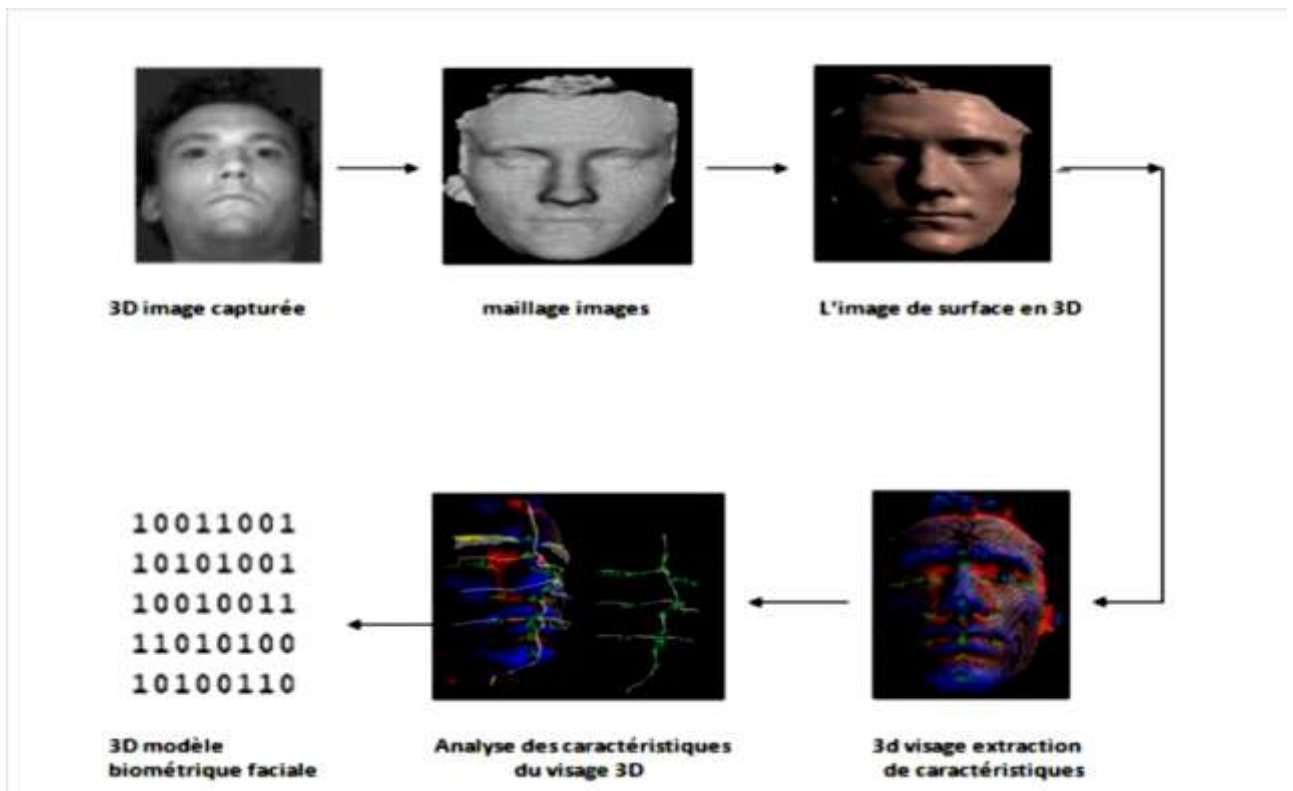


Figure 1.12. Processus de reconnaissance faciale 3D [4].

1.2.4 Reconnaissance de visage 3D

1.2.4.1 Techniques d'acquisition de visage 3D [4]

Durant la dernière décennie, les techniques de numérisation, de capture et de reconstruction de visages 3D sont témoins d'un formidable développement. Dans cette section, nous résumons trois

types de techniques : lumière structurée « structured light », multi-vues stéréo « multi-view stereo » et stéréo photométrique « photometric stereo ».

1) Lumière structurée (structured light) : En tant que technique de balayage "active", la lumière structurée a été largement utilisée dans l'acquisition de données de visages 3D.

La configuration de base de cette technique comprend une caméra et un projecteur de lumière dans les relations géométriques données. Le projecteur projette un ou plusieurs motifs de lumière codés sur la scène. La forme des signaux est extraite par la mesure de la déformation de motifs lumineux à la surface de l'objet. La figure 1.13 illustre un exemple des motifs de lumière codés et projetés sur un visage. Actuellement, un scanner 3D structuré à base de lumière, peut capturer des séquences statiques ou des scans du visage en 3D ainsi que des images 2D du visage en temps réel, ou même à haute vitesse. L'inconvénient est que les visages des utilisateurs devraient être limités dans une quantité limitée de la zone couverte non seulement par la lumière projetée structurée, mais aussi visible sur la caméra. Certains scanners 3D populaires basés sur des techniques d'éclairage structurés sont « Minolta Vivid 900/910 series » [19], « Inspeck Mega Capturor II 3D » [20], et « Kinect » [21], etc.



Figure 1.13. La projection d'une lumière structurée sur un visage.

2) Photométrie stéréo (Stereo photometric) : La photométrie stéréo, introduite à l'origine par Woodham en 1980, est une technique pour estimer les normales de la surface d'objets en capturant un ensemble d'images de l'objet sous différentes conditions d'éclairage. La figure 1.14 montre quelques reconstructions produites par quatre feux standards photométrie stéréo et la stéréo photométrie multispectrale. Cette technique est très sensible à la présence d'ombres, aux faits saillants et à l'éclairage non uniforme. Un système stéréo photométrique très récemment développé et utilisé pour l'acquisition de données de visages 3D est introduit par [22].

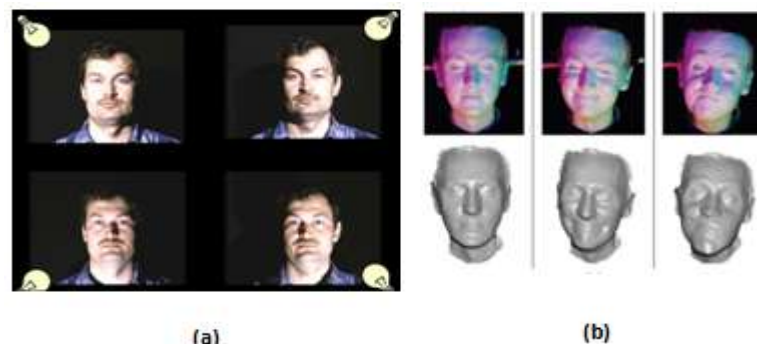


Figure 1.14. Comparaison de la stéréo photométrique (a) et stéréo photométrique multi spectrale (b).

3) Multi-vues stéréo (Multi-view stereo) : Comme une technique « passive » la capture multi-vue stéréo a également été largement utilisée dans l'acquisition de données faciales 3D. Cette technique utilise plusieurs caméras calibrées placées à différents points de vue du sujet. Sous diverses contraintes, les points correspondants de différentes images de bon sens peuvent être mis en place et ainsi la forme du visage peut être reconstruite. Cette technique ne nécessite pas de clignotant, puisque tout l'appareil peut enregistrer le même sens en même temps avec des lumières constantes. Cependant, la reconstruction précise des surfaces lisses (par exemple certaines régions du visage sans beaucoup de caractères de texture) est très difficile en utilisant cette technique.

En outre, la reconstruction 3D doit être utilisée hors ligne en raison de l'énorme complexité des calculs impliquant cette technique. Des systèmes commerciaux populaires qui emploient la technique de stéréo multi-vue sont le système de capture de visage dynamique DI3D et le système stéréo 3D dynamique 3DMD. Les deux systèmes ont été utilisés pour l'acquisition de bases de données de visage 3D. La configuration de base et des visages reconstruits par ce système sont représentés dans la figure 1.15.



Figure 1.15: Droite : systèmes de capture, Gauche : modèle de visage capturé en utilisant une configuration de studio de plusieurs caméras.

La reconnaissance générale de visage 3D fait référence à deux catégories d'applications. Une application appelée reconnaissance faciale 3D ou identification de visage 3D et une autre application appelée vérification de visage 3D ou authentification. L'identification de visage 3D est un problème de correspondance "un contre tous" alors que la vérification de visage 3D est un problème de correspondance "un contre un".

1.2.5 Base de données 3D de visage [4]

De nombreux organismes de recherche ont construit différentes bases de données de visage 3D pour évaluer les algorithmes de reconnaissance faciale 3D, par exemple : XM2VTS, 3D_RMA, MPI, York, Notre Dame, Gavab, BU-3DFE, BU-4DFE...etc. On peut citer des bases de données du visage les plus utilisées en 3D contenant des variations d'expression. Par ailleurs, chaque base de données faciale 3D a été conçue pour résoudre un ou plusieurs problèmes spécifiques de recherche dans ce domaine :

- Modèles d'un nombre important et variété démographique de sujets, comprennent des parties cachées de la face, et expressions faciales neutres, modèles d'une personne donnée à intervalles répétés de temps important : FRGC v1.0 [23].

- Modèles d'un grand nombre et variété démographique de sujets, modèles d'une personne donnée à intervalles répétés de temps : FRGC v2. [23].
- Modèles avec des variations importantes d'expressions faciales : BU-3DFE [24], Bosphorus [25], FRGCv2.0[23], CASIA [26], York [27], etc.
- Les modèles de jumeaux : 3DTEC [28].

Dans notre travail, une de ces bases de données a été choisie pour les évaluations et les analyses expérimentales de notre approche proposée, c'est FRGC v1.0. Les chapitres 4 et 5 expliciteront cette approche en utilisant cette base.

1.3 Conclusion

Dans ce chapitre, notre étude nous a permis de passer en revue les différentes technologies utilisées dans la biométrie : l'identification et la vérification. Nous avons aussi donné un aperçu de vu sur les techniques de mesure de leurs performances. Nous avons détaillé ce qu'est la biométrie en général et justifié le choix de la reconnaissance du visage en particulier.

Enfin, nous avons mis en évidence les différentes difficultés inhérentes à la reconnaissance automatique de visages (2D) et présenté ensuite la reconnaissance de visage 3D à travers la présentation des techniques 3D. Les techniques utilisées pour la reconnaissance de visage sont détaillées dans le chapitre suivant.

La reconnaissance faciale et ses différentes techniques

2.1 Introduction

L'utilisation des techniques de reconnaissance faciale a connu un développement accéléré, notamment avec l'utilisation efficace de nouvelles technologies, et principalement l'ordinateur et sa capacité sans cesse croissante dans le domaine du traitement d'images ainsi que celui de la fouille de données (data mining). Ceci leur a conféré de nouvelles applications et permis de développer davantage le domaine de la reconnaissance faciale.

Malgré les améliorations et la mise en œuvre de systèmes de reconnaissance de visages 2D qui ont un rendement acceptable dans des conditions contrôlées, des évaluations relativement récentes ont fait leur apparition comme la technique FRVT (Face Recognition Vendor Test) en 2002. Ceci a pu démontrer que la performance de ces systèmes se dégrade considérablement en fonction des conditions de pose, des paramètres liés à l'éclairage ainsi que des variations entre les conditions de prise de l'image de test et celles des images de la galerie. Depuis le début des années 1990, sur la base des progrès constants dans la technologie des capteurs 3D, le processus d'acquisition tridimensionnelle est devenu plus abordable en termes de coût, plus rapide et plus fiable. Dans cette optique, de nombreux chercheurs se sont tournés vers les scans 3D du visage comme une solution potentielle pour pallier les limites inhérentes à la reconnaissance de visage 2D [4].

Dans ce chapitre, nous présentons, dans un premier temps, les techniques de détection de visage. Nous nous proposons également de mettre à jour les méthodes de reconnaissance de visage qu'elles soient bidimensionnelles ou tridimensionnelles. Le but de ce chapitre est d'évaluer les capacités et les lacunes des algorithmes en cours, afin d'identifier les problèmes et les orientations pour la recherche future dans ce domaine.

2.2 Reconnaissance de visage 2D

Plusieurs méthodes d'identification de visages ont été proposées durant les vingt dernières années. L'identification de visage est un axe de recherche ouvert attirant des chercheurs venants de disciplines différentes : psychologie, reconnaissance de formes, réseaux neuraux, vision artificielle et infographie.

Avant de détailler les différentes techniques liées à la reconnaissance de visage 2D, nous allons d'abord présenter un aperçu des études faites par les chercheurs en cognition et en reconnaissance faciale du visage. En effet, la connaissance des résultats de ces études est importante, car elle permet le développement de nouvelles approches. Le but ultime de la

Chapitre 2 : La reconnaissance faciale et ses différentes techniques

reconnaissance faciale est de rivaliser, voir même dépasser, les capacités humaines de reconnaissance.

Une image du visage est un signal à 2 dimensions, acquis par un capteur digital (caméra numérique, scanner). Ce capteur codera la couleur ou l'intensité des différents points de l'image dans une matrice de pixels à deux dimensions. Après normalisation de l'image et mise à l'échelle vers une taille fixe (ex. : $m*n$), l'image du visage peut être considérée comme un vecteur dans un espace multidimensionnel RMN. Le nombre de points constituant cet espace devient rapidement très grand, même pour les images de petite dimension.

Pour les méthodes non paramétriques, le nombre d'exemples nécessaires afin de représenter efficacement la distribution des données peut être insuffisant. Le cas du visage est assez particulier. Le visage est formé par des surfaces lisses et une texture régulière. Il possède une forte symétrie (dans le cas des images frontales) et il est formé à partir des mêmes objets (yeux, nez et bouche), qui ont les mêmes positions pour tous les visages. Ces spécificités donnent lieu à plusieurs constatations :

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

En 1994, Ruderman a démontré que les images naturelles possèdent une grande redondance statistique [29]. En 1996, Penev [30] a démontré que dans le cas précis des images de visages normalisées, cette redondance statistique s'avère très forte.

L'appartenance à un même sous-espace et la redondance statistique qui caractérise les visages expliquent l'intérêt des chercheurs pour les méthodes de reconnaissance basées sur les approches par réduction d'espace, appelées également "les approches globales". Le fait que les visages soient confinés dans un sous-espace permet de définir une base (orthogonale, ou non orthogonale) dans cet espace. Cette base devrait représenter de façon plus pertinente n'importe quel point de l'espace des visages. L'avantage de la redondance statistique est qu'elle permet une extraction d'une structure simple des caractéristiques importantes et pertinentes de l'image du visage. Cette structure permettrait de représenter le visage tout en gardant l'information la plus importante et, par conséquent, de réduire la dimensionnalité de l'espace visage. Tout l'intérêt des approches globales est la construction de cette base de projection qui permettra de comparer, de reconnaître ou d'analyser l'information essentielle des visages [31].

Selon Albert Mehrabian [32], l'expression faciale d'un interlocuteur compte 55 % de l'effet d'un message oral et l'intonation de la voix contribue à 38 % tandis que la partie verbale a seulement 7 % d'effet. De ceci s'inspirent plusieurs travaux qui sont orientés vers l'étude des expressions faciales notamment la reconnaissance et la reconstruction.

La littérature présente un nombre important de techniques qui traitent la reconnaissance des expressions faciales, toutefois nous nous intéressons seulement à celles qui sont toujours d'actualité dans les travaux de recherches du domaine. Avant d'aller plus loin, il est essentiel de noter que

plusieurs techniques traitent à la fois la détection de visage et ses composants et la reconnaissance du visage et des expressions faciales.

L'article de Tana et al. [33] donne un état de l'art récent et complet sur les dernières techniques développées. Dans cette section, nous présenterons les approches 2D de la reconnaissance de visage les plus connues. Ces dernières peuvent être subdivisées en trois catégories : les approches holistiques ou globales, les approches locales et les approches hybrides.

2.2.1 Approches globales

Ces méthodes identifient un visage en utilisant l'image entière de ce dernier comme entrée du système de reconnaissance. Chaque image du visage de dimension (n, m) est représentée par un vecteur simple de dimension $n \times m$, en concaténant les valeurs du niveau de gris de tous les pixels de l'image du visage. Par exemple, considérons une petite image de 64×64 , en niveaux de gris codés sur 8 bits (donc de 0 à 255). L'espace I contenant tous les vecteurs images de visages est appelé espace images. L'avantage de cette représentation réside dans le constat qu'elle préserve implicitement les informations de texture et de forme nécessaires pour la reconnaissance de visages. De plus, elle permet une meilleure capture de l'aspect global du visage que les représentations locales. Toutefois, son utilisation de techniques de modélisation de sous-espace a fait avancer la technologie de reconnaissance faciale de manière significative. Nous pouvons distinguer deux types de techniques parmi les méthodes globales : les techniques linéaires et les techniques non linéaires [34].

2.2.1.1 Les techniques linéaires

Les techniques linéaires projettent linéairement les données d'un espace de grande dimension (par exemple, l'espace de l'image originale) sur un sous-espace de dimension inférieure. Malheureusement, ces techniques sont incapables de préserver les variations non convexes des variétés (géométriques au sens mathématique du terme) de visages afin de différencier des individus. Dans un sous-espace linéaire, les distances euclidiennes et plus généralement les distances de Mahalanobis, qui sont normalement utilisées pour comparer des vecteurs de données, ne permettent pas une bonne classification entre les classes de formes "visage" et "non-visage" et entre les individus eux-mêmes. Ce facteur crucial limite le pouvoir des techniques linéaires pour obtenir une détection et une reconnaissance du visage très précises. [4]

Une méthode très populaire, basée sur la technique ACP, est la méthode Eigenface [35]. La méthode Eigenface se base sur l'idée qu'une image de visage x peut être traitée (compressée ou reconstruite) à partir d'une image moyenne m , à laquelle on ajoute un petit nombre d'images u_i de taille bien déterminée.

$$\tilde{x} = m + \sum_{i=1}^{X-1} (a_i u_i) \quad (1)$$

Pour avoir les images u_i , il suffit d'appliquer une PCA sur les images d'apprentissage, pour ce faire il est essentiel de convertir chaque image en un vecteur de données (figure 1).

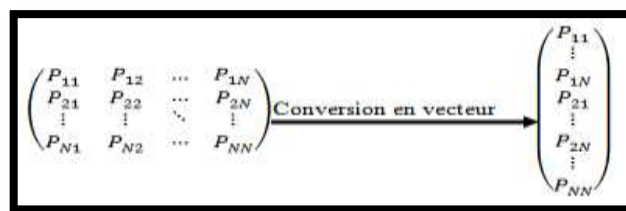


Figure 2.1. Conversion d’image en vecteur colonne.

Une image $N \times N$ est transformée en un vecteur de dimension N^2 . Puisque, la comparaison dans l’espace N^2 n’est pas performante, la méthode calcule la moyenne des images puis les normalise par rapport à cette moyenne. Ensuite, elle génère la matrice de covariance qui présente les corrélations et les dépendances entre les dimensions de l’ensemble. Le calcul de la matrice de covariance résultat donne N^2 vecteurs propres ; ce qui est assez grand.

À partir de cette matrice, on maintient les X premiers vecteurs propres qui sont dans un ordre décroissant des valeurs propres associées. Dans la plupart des cas X est compris dans l’intervalle $[\frac{3}{4}M, M]$. Une transformation inverse à celle appliquée antérieurement est exécutée sur les vecteurs retenus pour donner une nouvelle série d’images dites eigenfaces.

Enfin, on calcule la matrice des poids des visages de notre base initiale en les projetant dans un sous-espace engendré par les visages propres (figure 3).

$$\begin{bmatrix} P_{1[11]} & \dots & P_{1[1N]} & P_{1[21]} & \dots & P_{1[2N]} & \dots & P_{1[NN]} \\ P_{2[11]} & \dots & P_{2[1N]} & P_{2[21]} & \dots & P_{2[2N]} & \dots & P_{2[NN]} \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ P_{M[11]} & \dots & P_{M[1N]} & P_{M[21]} & \dots & P_{M[2N]} & \dots & P_{M[NN]} \end{bmatrix} \times \begin{bmatrix} P_{1[11]} & P_{2[11]} & \dots & P_{X[11]} \\ P_{1[1N]} & P_{2[1N]} & \dots & P_{X[1N]} \\ P_{1[21]} & P_{2[21]} & \dots & P_{X[21]} \\ P_{1[2N]} & P_{2[2N]} & \dots & P_{X[2N]} \\ P_{1[NN]} & P_{2[NN]} & \dots & P_{X[NN]} \end{bmatrix} = \begin{bmatrix} P_{11} & \dots & P_{1X} \\ \vdots & \ddots & \vdots \\ P_{M1} & \dots & P_{MX} \end{bmatrix}$$

Matrice d’images en ligne x (sous-Matrice X-eigenfaces)^T = Matrice de poids

Figure 2.2. Calcul de la matrice de poids.

Pendant la phase de test d’une nouvelle image, les mêmes étapes sont appliquées pour créer le visage propre et calculer les poids de cette dernière image dans notre espace de visages propres. A la fin nous déterminons l’identité de l’expression faciale présente dans la nouvelle image en comparant ses poids par rapport aux éléments de la matrice des poids dans la partie caractérisation (voir chapitre 3).

Selon Richard Szeliski [36], la reconnaissance de visage (respectivement des expressions faciales) par eigenfaces possède un grand avantage à savoir la réduction de la dimension et de la sensibilité au bruit. Cependant, elle n’est pas optimale lors de séparation de classes en cas de présence d’une large variabilité de lumière au sein d’une même classe associée à une variabilité interclasses. Dans notre cas de figure, ce problème ne sera pas présent.



Figure 2.3. Exemples des 5 premiers visages propres construits avec l'approche ACP.

L'Analyse en composantes indépendantes (ou ICA pour Independent Component Analysis) a été introduite par Héroult, Jutten et Ans dans [37, 38]. Elle devint populaire lors de son utilisation dans le domaine du traitement du signal pour la séparation de sources aveugles dans [39] et [40]. Le but est d'exprimer un ensemble de n variables aléatoires x_1, \dots, x_n comme une combinaison linéaire de n variables aléatoires statistiquement indépendantes s_j :

$$x_j = a_{j,1}s_1 + \dots + a_{j,n}s_n \quad \forall j \quad (2)$$

Ou sous forme matricielle, avec :

$$x = [x_1, \dots, x_n]^T, \quad s = [s_1, \dots, s_n]^T \quad \text{et } A \text{ est une matrice contenant les coefficients } a_{ij}.$$

Le but de l'analyse en composantes indépendantes est l'estimation des composantes originales s_i , ou de manière équivalente des coefficients a_{ij} . Par définition, les variables aléatoires s_i sont mutuellement indépendantes et la matrice de mélange est donc inversible.

Bartlett et al [41] ont montré que les statistiques de premier et de second ordre permettent d'extraire des informations uniquement sur le spectre d'une image d'amplitude tout en rejetant le spectre de phase.

Cependant, certaines expériences ont mis en évidence que la capacité humaine à reconnaître des objets peut être due au spectre de phase. C'est aussi la raison pour laquelle ils ont étudié la technique ICA comme un algorithme puissant pour la reconnaissance faciale. De plus, la technique ICA sépare les moments d'ordre élevé de l'image d'entrée en plus des moments de second ordre utilisés par la technique PCA. Cela peut être considéré comme une généralisation de la technique PCA.

Le Multidimensional Scaling (MDS) est une autre méthode très connue de réduction de dimension linéaire [42]. Au lieu de conserver la variance des données au cours de la projection, elle s'efforce de préserver toutes les distances entre chaque paire d'exemples $\text{dist}(x_i, x_j)$ en cherchant une transformation linéaire qui minimise l'énergie :

$$\varepsilon_{\text{mds}} = \sum_{i,j=1}^n (\text{dist}(x_i, x_j) - \|u_i - u_j\|)^2 \quad (3)$$

Une autre méthode qui vise à représenter le visage sans employer la notion de classe est la factorisation non négative des matrices (Non Negative Matrix Factorization, NMF) [43]. L'algorithme de la NMF, comme celui de PCA, représente le visage comme combinaison linéaire de vecteurs de la base de l'espace réduit. La différence est que la NMF n'autorise pas les éléments

négatifs dans les vecteurs de la base dans les poids de la combinaison. Au final, certains vecteurs de l'espace réduit par PCA (visages propres) ressemblent à des versions distordues du visage entier alors que ceux réduits par la NMF sont des objets localisés qui correspondent mieux à la notion de parties du visage.

La présente décennie, plusieurs chercheurs ont développé diverses extensions de la méthode eigenface, comme l'eigenface probabiliste [44], les méthodes basées sur l'Analyse Discriminante Linéaire (ADL) [45, 46], ... etc. Ces approches disposent des performances supérieures à celles qui sont données par la méthode PCA. Tandis que la technique PCA construit un sous-espace pour représenter de manière "optimale" pour seulement "l'objet" visage, la technique LDA construit un sous-espace discriminant pour distinguer de façon "optimale" les visages de différentes personnes. Appelée également "Fisher Linear Discriminant", la LDA est parmi l'une des méthodes les plus utilisées pour la reconnaissance de visage. Elle utilise le critère de réduction qui se base sur la notion de séparabilité des données par classe qui date de 1936 [47]. Cette méthode fut appliquée aux visages en 1996 [48].

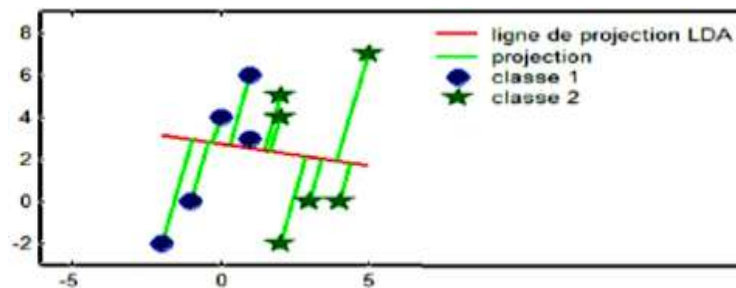


Figure 2.4. Projection par LDA avec séparation entre les classes.

Cependant, ceci n'est plus vrai si un seul exemple d'apprentissage par personne est disponible, en raison du fait que la plupart de ces techniques vont, soit se réduire à l'approche eigenface basique, ou bien ne fonctionnent plus. Ces méthodes se résument en :

- Les **méthodes basées sur l'Analyse Discriminante Linéaire (ADL)** : déterminent les directions de projection les plus discriminantes dans l'eigenspace. Pour cela, elles maximisent les variations inter personne par rapport aux variations intra-personne.
- La **méthode probabiliste** : transforme le problème d'identification de visage en un problème de classification en deux classes. Elle évalue la probabilité de la différence entre une image de test et une image prototype appartenant aux classes intra-personne et inter personne. Notons que la distribution intra-personne ne peut pas être évaluée dans le cas d'un exemple par personne, et la méthode se ramène aussi à la méthode eigenface [44].

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

2.2.1.2 Les techniques non-linéaires

L'idée est d'utiliser une fonction noyau pour construire un espace de grande dimension dans lequel le problème devient linéaire. On peut ainsi appliquer des méthodes linéaires de réduction de dimension lorsque la structure intrinsèque des données n'est pas linéaire. Ces méthodes utilisent généralement « l'astuce du noyau » (de l'anglais kernel trick) qui établit que tout algorithme formulé avec une fonction noyau peut être reformulé avec une autre fonction noyau. L'Analyse en Composantes Principales à Noyaux (Kernel Principal Component Analysis, KPCA) est la reformulation non linéaire de la technique linéaire classique qu'est l'analyse en composantes principales en utilisant des fonctions à noyaux [49]. Ainsi, depuis plusieurs années, la reformulation de techniques classiques à l'aide de « l'astuce du noyau » a permis l'émergence de nombreuses techniques comme les machines à support de vecteurs (Support Vector Machine, SVM) [50]. La technique PCA à noyaux calcule les principaux vecteurs propres de la matrice de noyaux plutôt que la matrice de covariance [4].

La technique PCA à noyaux calcule d'abord la matrice du noyau K des points x_i dont les entrées sont définies par [50] :

$$k_{i,j} = k(x_i, x_j) \quad (4)$$

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

$$k_{i,j} = k_{i,j} - \frac{1}{n} \sum_l k_{i,l} - \frac{1}{n^2} \sum_l k_{j,l} + \frac{1}{n} \sum_{lm} k_{lm} \quad (5)$$

L'utilisation de ces méthodes de projection de l'espace des images sur l'espace de caractéristiques est non linéaire et permet ainsi dans une certaine mesure de réduire la dimension des images de façon meilleure. Cependant, bien que ces méthodes permettent souvent l'amélioration des taux de reconnaissance sur des jeux de tests donnés, elles sont trop flexibles pour être robustes à de nouvelles données, contrairement aux méthodes linéaires [4].

2.2.2 Approches locales

Celles-ci utilisent les caractéristiques faciales locales pour la reconnaissance de visage. Nous pouvons classer les méthodes locales en deux catégories : celles basées sur les caractéristiques locales : extractions et localisation des points caractéristiques, et celles basées sur les apparences locales : partitions des images de visage en régions caractéristiques [33].

2.2.2.1 Méthodes basées sur les caractéristiques locales : [8]

Ces méthodes sont efficaces. Cependant, leurs performances dépendent particulièrement de la précision de localisation des points caractéristiques. La tâche reste très difficile en pratique, plus précisément dans des situations où la forme et l'apparence du visage peuvent fortement changer. Les approches basées sur l'extraction de points caractéristiques peuvent être subdivisées en deux catégories : les approches géométriques et les approches basées sur les graphes.

- ❖ *Approches géométriques* : Elles sont basées sur l'extraction de la position relative des éléments qui constituent le visage (tels que le nez, la bouche et les yeux). La plupart des

approches géométriques utilisent des points d'intérêt (comme les coins de la bouche et des yeux). Le coût de stockage des techniques géométriques est très bas comparé à celui des autres techniques. Toutefois, les approches purement géométriques présentent quelques inconvénients, notamment :

- Les caractéristiques géométriques sont généralement difficiles à extraire, surtout dans des cas complexes : illumination variable, occultations...
 - Les caractéristiques géométriques seules ne suffisent pas pour représenter un visage, tandis que d'autres informations utiles comme les niveaux de gris de l'image ne sont pas du tout exploitées.
- ❖ *Approches basées sur les graphes* : Au lieu d'utiliser des méthodes purement géométriques, certains chercheurs ont choisi de représenter les caractéristiques locales du visage sous forme de graphes. La reconnaissance de visages est alors formulée comme un problème de mise en correspondance de graphes. Cependant, une fois construit, le graphe topologique ne peut pas être modifié. Or, les images du visage changent facilement d'apparence en raison des différentes variations (illumination, expression, pose, etc.), et du coup un schéma de graphe topologique fixe n'est plus adéquat.

Wiskott et al. [51] ont étendu l'utilisation de LDA à une méthode très connue appelée Elastic Bunch Graph Matching (EBGM), où les nœuds des graphes sont situés sur un certain nombre de points sélectionnés du visage. Wiskott et al. ont utilisé les ondelettes de Gabor pour extraire les caractéristiques des points détectés, car les filtres de Gabor sont robustes aux changements d'illumination, aux distorsions et aux variations d'échelle (voir les détails sur les filtres de Gabor dans le chapitre 3). De ce fait, l'EBGM fut l'un des algorithmes les plus performants lors de la compétition de FERET en 1996.

Pour conclure, de nombreuses méthodes basées sur l'extraction de points caractéristiques ont été proposées et ces méthodes peuvent être utiles et efficaces pour la reconnaissance de visages dans le cas où une seule image de référence est disponible. Cependant, leurs performances dépendent beaucoup de l'efficacité des algorithmes de localisation des points caractéristiques du visage. [4]

En pratique, la tâche de détection précise des points caractéristiques n'est pas facile et n'a pas été résolue de manière fiable, en particulier dans les cas où la forme ou l'apparence d'une image du visage peuvent beaucoup varier [52]. Le point faible de ces méthodes reste le temps de calcul des paramètres et la complexité de la comparaison entre référence et test.

2.2.2.2 Les méthodes locales basées sur l'apparence du visage

Une fois les régions locales définies, il s'agit de choisir la meilleure manière de représenter les informations de chaque région. Cette étape est critique pour les performances du système de reconnaissance. Les caractéristiques couramment utilisées sont : les coefficients de Gabor [51,53,54], les ondelettes de Harr [55], les transformées de Fourier, SIFT (Scale Invariant Feature Transform) [56], les caractéristiques basées sur la méthode LBP (Local Binary Pattern) [57], LPQ (Local Phase Quantization) [58], WLD (Weber Law Descriptor) [59], BSIF (Binarized Statistical Image Features) [60]. Dans cette section on discutera en détail les méthodes choisies à l'utiliser dans notre approche. [4]

❖ SIFT et sa variante

La technique Scale-Invariant Feature Transform (SIFT), que l'on peut traduire par « transformation de caractéristiques visuelles invariantes à l'échelle », est une approche pour la détection et l'extraction de descripteurs de caractéristiques locales qui sont raisonnablement invariantes aux changements dans l'éclairage, le bruit d'image, la rotation, la mise à l'échelle et de petits changements de points de vue. Il a été développé en 1999 par le chercheur David Lowe [61]. Selon [61], le fonctionnement de l'algorithme SIFT comporte cinq étapes :

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

La première étape de l'algorithme est la détection des points d'intérêt, dits points clés. Un point-clé (x,y,σ) est défini d'une part par ses coordonnées sur l'image (x et y) et d'autre part par son facteur d'échelle caractéristique (σ). En toute rigueur, il s'agit d'une zone d'intérêt circulaire, le rayon de la zone étant proportionnel au facteur d'échelle. Il s'ensuit une étape de re-convergence et de filtrage qui permet d'améliorer la précision sur la localisation des points-clés et d'en éliminer un certain nombre d'entre eux jugés non pertinents. Chaque point-clé restant est ensuite associé à une orientation intrinsèque, c'est-à-dire ne dépendant que du contenu local de l'image autour du point clé, au facteur d'échelle considéré. Elle permet d'assurer l'invariance de la méthode à la rotation et est utilisée comme référence dans le calcul du descripteur, qui constitue la dernière étape de ce processus [62].

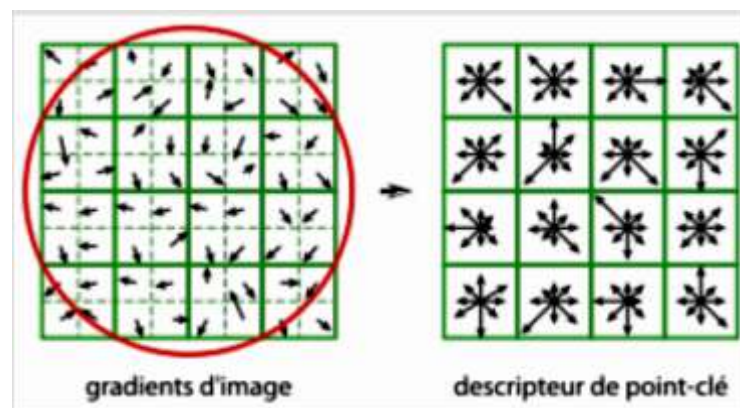


Figure 2.5. Construction d'un descripteur SIFT.

Outre la reconnaissance d'objet, les avantageuses propriétés des descripteurs SIFT (caractère discriminant, invariance à la translation, à la rotation et au changement d'échelle et robustesse aux transformations affines en général (distorsions), aux changements de points de vue 3D ainsi qu'aux changements de luminosité) en font un excellent choix pour d'autres applications ; localisation de robot en environnement inconnu, l'assemblage de panoramas, modélisation, reconnaissance et suivi d'objets 3d, descripteurs SIFT 3d pour la reconnaissance de mouvements humains,.....

Les variantes et extensions principales de SIFT on peut citer :

RIFT, G-RIF Generalized robust invariant feature), SURF (Speeded Up Robust Features), PCA-SIFT et GLOH Gradient location-orientation histogram), et bien d'autres variantes que nous ne citerons pas.

❖ LBP et sa variante récente [63, 4]

La méthode LBP (Local Binary Pattern) originale étiquette les pixels d'une image avec des nombres décimaux, qui sont appelés LBP qui codent la structure locale autour de chaque pixel. On procède ainsi, comme illustré sur la figure 7 : chaque pixel est comparé à ses huit voisins dans un voisinage 3×3 en soustrayant la valeur du pixel central ; les valeurs strictement négatives résultantes sont codées avec 0 et les autres avec un. Pour chaque pixel donné, un nombre binaire est obtenu en concaténant l'ensemble de ces valeurs binaires dans le sens horaire, ce qui commence à partir de l'une de ses voisines en haut à gauche.

La valeur décimale correspondante du nombre binaire généré est ensuite utilisée pour marquer le pixel donné. Les nombres binaires dérivés sont appelés codes de LBP [64]. Jin et al. [65] ont amélioré l'opérateur de LBP, ils ont proposé une LBP améliorée (Improved LBP, ILBP).

Yang et Wang [66] ont proposé Hamming LBP pour améliorer la capacité discriminante de la LBP originale. Plus récemment, Guo et al. [67] ont proposé CLBP (Complete LBP), qui, à notre avis, est tout à fait similaire avec ELBP. SLBP (Soft LBP) a été introduite dans [68] qui emploie deux fonctions d'appartenance floues au lieu de l'indicateur du signe de LBP.

Zhang et al. [69] proposent un nouvel ensemble des traits distinctifs rectangulaires pour la détection de visage sous le nom de multiples blocs de modèles binaires locaux abrégé en MB-LBP (Multi-block Local Binary Pattern). Les MB-LBP codent les intensités des régions rectangulaires par le modèle binaire local. Les modèles résultats décrivent plusieurs structures locales dans l'image et arrivent à capturer des structures avec différentes tailles.

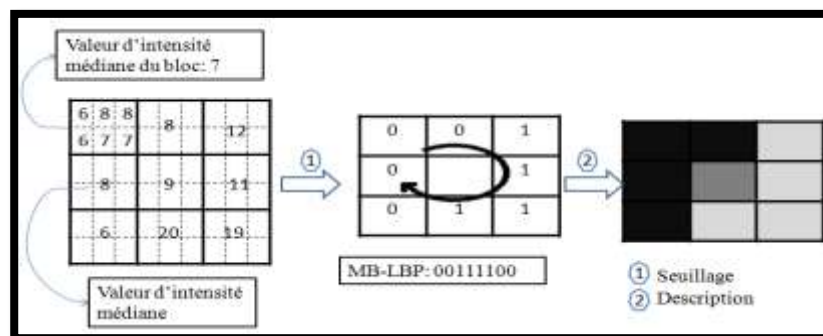


Figure 2.6. Les caractéristiques MB-LBP proposées par Zhang et al. [69].

❖ Méthode de Viola et Jones

Nous nous intéressons particulièrement à décrire de manière brève certains travaux existants à partir des années 2000. La célèbre technique de détection de visage dans la dernière décennie est celle proposée par Paul Viola et Michael Jones en 2001 [70,71]. Cette méthode combine quatre principes clés qui sont les caractéristiques rectangulaires simples appelées des caractéristiques pseudo-Haar en raison de leur similitude avec les ondelettes de Haar, l'approche d'image intégrale pour la détection rapide et efficace des caractéristiques, la méthode d'apprentissage adaptative AdaBoost

Chapitre 2 : La reconnaissance faciale et ses différentes techniques

[72] (publiés par Freund et Schapire en 1996) cherchant à minimiser l'erreur de classification, et l'algorithme en cascade de classifieurs : c'est une chaîne qui combine plusieurs classifieurs par étages triés par ordre croissant de complexité. (Pour plus de détails, voir le chapitre 3).

Dans la suite, nous allons détailler ces étapes clés tout en précisant l'apport engendré de chacune d'entre elles sur la performance et l'efficacité de la méthode. Nous finirons notre description à propos la méthode « Viola&Jones » par l'énumération des avantages ajoutés par chacune des contributions citées :

- Les caractéristiques pseudo-Haar permettent la détection des objets selon plusieurs échelles (détection multi-échelles).
- L'image intégrale permet le calcul des caractéristiques en temps réel.
- L'algorithme AdaBoost sélectionne les caractéristiques les plus discriminantes pour la classification et forme un classifieur de bonne performance.
- La cascade minimise le temps de calcul et affine les frontières de classification.

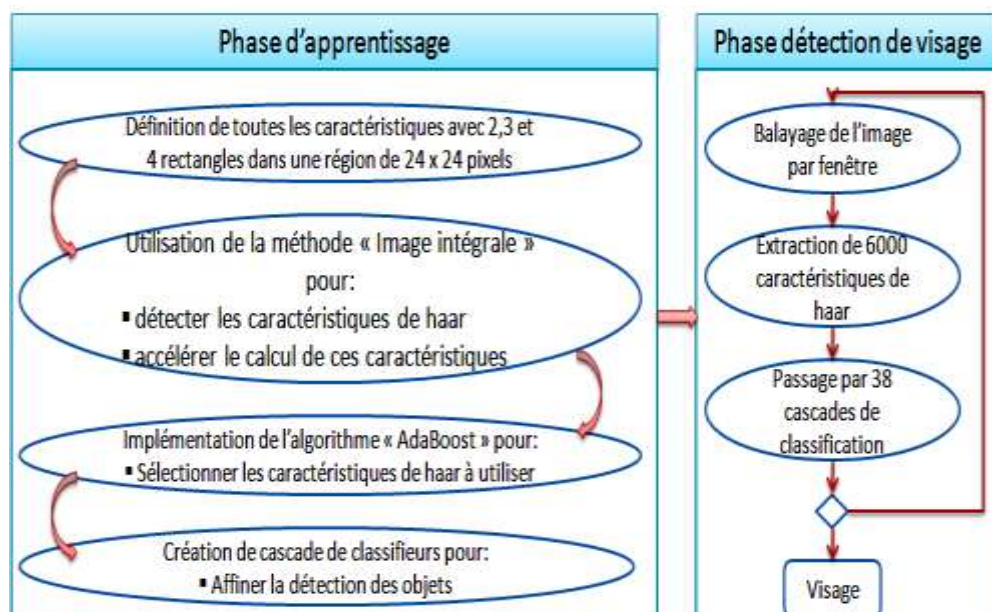


Figure 2.7. La chaîne d'exécution de la méthode « Viola&Jones ».

Plusieurs chercheurs se sont inspirés de l'approche « Viola&Jones », tout en proposant des modifications au niveau des caractéristiques d'extraction pseudo-Haar et/ou au niveau de l'utilisation de l'algorithme de dopage afin de perfectionner la détection faciale.

Les auteurs de [73] présentent également une procédure de post-optimisation d'une cascade dopée. Le taux d'amélioration de performance mené par la première intervention par rapport à « Viola&Jones » est égal à 10 % et celui engendré par la deuxième intervention est environ 12.5 %. Leur combinaison atteint un taux de progression supérieur à 23 %.

Chapitre 2 : La reconnaissance faciale et ses différentes techniques

En 2003, Viola et Jones améliorent leur méthode initiale pour faire face au problème de détection de plusieurs vues de visage en même temps. Ils proposent dans [74] des détecteurs pour différentes vues de visage (face et profil) et ils effectuent l'apprentissage d'un arbre de décision pour déterminer automatiquement le détecteur adéquat au visage présent dans l'image. Ces créateurs montrent bien que les modifications réalisées donnent de bons résultats et maintiennent la vitesse d'exécution de leur première approche.

Dans [69], les auteurs utilisent une méthode d'apprentissage basée sur l'algorithme Gentle AdaBoost pour la sélection des caractéristiques convenables et la construction des classificateurs d'une cascade détectrice. Selon Zhang et *al*, les MB-LBP atteignent un taux de réussite au voisinage de 15 % de plus à celui obtenu par la méthode « Viola&Jones ».

Dans [75], les chercheurs annoncent que leur système présente un taux de performance élevé par rapport à la méthode « Viola&Jones » et ceci en utilisant des filtres gaussiens à la place des caractéristiques pseudo-Haar de base. Meynet et *al*. mentionnent que ces filtres modélisent bien l'apparence de visage et ils sont simples à calculer et à créer. Une caractéristique est formée par une combinaison entre un filtre dans une première direction et sa dérivée première dans une direction orthogonale.

Le détecteur de Li et *al*. se compose de trois niveaux : le premier fonctionne sur les visages de toutes poses, le second ne détecte que des visages avec vue de profil gauche entre -90° et -30° , vue de face dans l'intervalle $[-30^\circ, 30^\circ]$ et vue de profil droite entre 30° et 90° . Le troisième niveau détecte des visages qui présentent des rotations avec sept angles plus petits. Les modifications apportées par [76] corrigent la non-détection de plusieurs vues en même temps de la méthode « Viola&Jones » et augmentent la vitesse de test avec un taux compris entre 1 % et 7 %.

Étant donné que nous traitons des séquences vidéo, la phase de détection implique implicitement la phase de suivi du visage dans la scène, puisque nous traitons la séquence vidéo image par image. Cette étape se décompose de trois tâches, à savoir, la détection du visage à l'aide de la méthode « Viola&Jones », l'extraction des traits faciaux (sourcils, yeux, nez, bouche, ...) et le suivi des déplacements du visage dans la scène.

2.2.3 Autres approches [4]

D'autres approches basées sur des modèles statistiques ont également été largement explorées. Un modèle statistique est une formalisation des relations entre les variables sous la forme d'équations mathématiques qui décrivent comment une ou plusieurs variables aléatoires sont liées à une ou plusieurs variables aléatoires. Le modèle est considéré comme statistique lorsque les variables ne sont pas déterministes, mais stochastiquement liées. Pour formuler l'analyse du visage avec le modèle statistique, nous construisons des modèles de l'apparence du visage et de ses variations et apprendre les manières dont la forme et la texture du visage humain varient dans une série d'images.

Des exemples des approches basées sur des modèles statistiques comprennent le modèle actif d'apparence (Active Appearance Model, AAM) [77], GMM (Gaussian Mixture Model) [78,79], 1D HMM (Hidden Markov Model) [80], pseudo-2D HMM [81], etc.

2.2.4 Discussion [4]

En comparant ces approches, les méthodes globales ont tendance à être moins sensibles au bruit. Cependant, les méthodes locales ont montré leur performance prometteuse au cours des dernières années.

Malgré que les techniques basées sur les caractéristiques locales permettent d'extraire des points caractéristiques locaux, il reste généralement quelques caractéristiques invariantes, même en présence de variations d'expression du visage ou des occlusions, la reconnaissance peut encore être faite en faisant correspondre ces caractéristiques invariantes. Par conséquent, les méthodes basées sur les caractéristiques locales sont potentiellement plus robustes que celles holistiques à des changements d'expression et des occlusions faciales.

On pourrait poser la question : qui est le meilleur descripteur local pour la reconnaissance faciale chez LBP, SIFT ? Luo et al. [82] ont montré que SIFT n'est pas aussi robuste que LBP à l'éclairage des effets sur l'ensemble de données FERET.

Ruiz-Del-Solar et al. [83] ont largement évalué ces méthodes, non seulement sur des bases de données comme FERET et FRGC, mais aussi sur la base de données UCH Face HRI conçue pour l'interaction homme-robot et l'ensemble de données LFW capturés dans des environnements contraignants. En ce qui concerne la robustesse aux variations d'éclairage, leur étude a montré que la technique LBP atteint la meilleure performance que la méthode SIFT sur la base de données FERET.

Il est difficile de tirer une conclusion générale pour classer ces descripteurs locaux, car leur rendement est déterminé par un certain nombre de facteurs complexes, lequel devrait être sélectionné pour la reconnaissance des visages qui dépend des exigences des applications.

2.3 Reconnaissance de visage 3D

Nous avons étudié des méthodes bidimensionnelles de reconnaissance des visages. Les études menées ces dernières années par les chercheurs en matière de reconnaissance 2D de visages robustes aux différents facteurs de variabilité (éclairage, pose, occultation) sont loin d'avoir atteint leurs objectifs. La reconnaissance de visages 3D constitue aujourd'hui une alternative prometteuse pour élucider ces problèmes, ceci dès l'apparition de dispositifs d'acquisition 3D performants.

L'avantage principal des approches basées sur les modèles 3D réside dans le fait que les modèles 3D conservent toutes les informations sur la géométrie du visage, ce qui permet d'avoir une représentation réelle de ce dernier. Semblables à la taxonomie de reconnaissance faciale 2D, les approches de reconnaissance de visage 3D peuvent également être classées en deux approches : globales et locales.

Beaucoup de recherches antérieures avaient accès à de très petites bases de données et, par conséquent, elles signalent des taux de reconnaissance étonnamment élevés. Au fil du temps, les enjeux sont devenus plus grands et les algorithmes ont produit des taux plus précis [8].

2.3.1 Approches globales [4]

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

2.3.1.1 *Iterative Closest Point (ICP)*

Dans ce cas, le problème de la reconnaissance 3D de visages est celui de l'alignement de deux surfaces 3D qui modélisent les deux visages à apparier. L'algorithme généralement utilisé est l'algorithme du plus proche voisin itéré, ou ICP (Iterative Closest Point), qui a été introduit par [84]. Il consiste en une optimisation alternée d'appariements et de transformations.

Medioni et Waupotitsch [85] ont effectué une reconnaissance de visage 3D en utilisant ICP pour la correspondance des surfaces faciales. Alors que la plupart des travaux utilisent des formes 3D acquises par un capteur à lumière structurée (structured-light). Wang et al. [86] ont proposé l'ICP partielle qui sélectionne une partie de paires de points de proximité pour calculer la mesure de similarité lors de l'enregistrement des surfaces du visage et afin de réduire les effets négatifs causés par les changements d'expression du visage.

Dans la plupart des travaux récents, l'ICP est largement présentée comme une étape clé dans la phase de prétraitement pour corriger les variations de pose 3D comme dans [87, 88]. Mohammadzade et Hatzinakos [89] ont proposé la méthode "Iterative Closest Normal Point" pour trouver les points correspondants entre une face de référence générique et chaque face d'entrée.

2.3.1.2 *Le sous-espace (PCA et LDA)*

Les méthodes basées sur le sous-espace ont été largement utilisées dans la reconnaissance faciale 2D. Plus récemment, ces techniques ont été aussi étendues aux données 3D de visage. Achermann et al. [90] ont appliqué l'approche du visage propre à la reconnaissance faciale 3D. Ils présentent les résultats d'un ensemble de données de 24 sujets qui ont chacun 10 images et trouvent un taux de reconnaissance de 100 %.

Heseltine et al. [91] évaluèrent l'approche de visage propre avec différentes distances comme la mesure de similarité pour la vérification du visage sur une grande base de données contenant 330 images de plus de 100 personnes. Ils conclurent que le plus faible taux d'erreur égalité (TEE) 17,8 % était atteint pour la distance de Mahalanobis et non avec les distances euclidiennes et cosinus.

Heseltine et al. [92] élargirent l'ensemble de tests à 1470 images de profondeur de 230 personnes et exploitèrent la LDA dans le même cadre. Ils affirmèrent que la LDA est plus performante que la PCA avec la distance cosinus. Avec la LDA, ils obtinrent un TEE= 15,3 %.

2.3.1.3 Distance de Hausdorff

La distance de Hausdorff est une mesure de la similarité de deux ensembles de points [93]. Elle est très générale et peut être appliquée à une grande variété de problèmes. La distance non dirigée de Hausdorff entre deux ensembles de points A et B est définie comme :

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (6)$$

Où $h(A, B)$ désigne la distance de Hausdorff dirigée, la distance de Hausdorff a été utilisée à l'origine pour des ensembles de points dans le plan 2D [93].

Achermann et Bunker [94] utilisent une extension de la distance de Hausdorff pour la reconnaissance de visage 3D. Ils rendent compte des expériences utilisant 240 images de profondeur, 10 images de chacune des 24 personnes et atteignent la reconnaissance de 100 % pour certains cas.

Pan et al. [95] ont comparé une approche de la distance de Hausdorff et celle fondée sur la technique PCA en reconnaissance faciale 3D. Dans les expériences avec des images de la base de données 3D_RMA, ils signalent un TEE de l'ordre de 3 % à 5 % pour l'approche de la distance de Hausdorff et 5 % à 7 % pour l'approche fondée sur la PCA. Russ et al. [96] ont exploré la correspondance sur les images 3D du visage.

2.3.1.4 Modèle déformable

Le modèle déformable est proposé afin de mieux décrire les transformations non rigides des visages. En utilisant des échantillons représentatifs dans la base d'apprentissage, le modèle déformable peut être utilisé pour [69] :

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

Lu et Jain [97, 98] proposent une autre méthode de reconnaissance faciale 3D en utilisant une approche de reconnaissance à base de ICP [99] qui est robuste par rapport aux variations d'expression du visage. Basée sur l'information apprise en utilisant le groupe de contrôle pour chaque sujet, l'approche peut générer différents modèles de visage d'expression neutre dans l'ensemble de la galerie.

Al-Osaimi et al. [100] ont pris les modes de déformations d'expression à partir des données de formation en vecteurs propres de PCA et ces modes sont ensuite utilisés pour transformer les déformations d'expression. Des mesures de similarité sont extraites en faisant correspondre les visages 3D transformés. L'approche est appliquée sur la base de données FRGC v2.

2.3.2 Approches locales [4]

Les approches locales ont été étudiées dans la reconnaissance faciale 3D puisqu'elles atteignent généralement de meilleurs résultats par rapport les approches globales. Ces approches sont classées dans [101] en cinq classes.

2.3.2.1 Les méthodes basées sur les descripteurs

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

Gupta et al. [102] proposent d'utiliser les ondelettes complexes à similarité structurelle métrique (Complex Wavelet Structural Similarity Metric, CW-SSIM) extraites des images de profondeur pour le visage pour la reconnaissance faciale 3D. La CWSSIM calcule efficacement et d'une façon robuste aux petites distorsions géométriques.

Berritti et Del Bimbo ont employé la détection des points-clés et la description de modèle SIFT pour identifier et caractériser les points-clés sur la représentation de l'image de gamme d'un visage. Ils ont proposé une approche de reconnaissance faciale 3D qui est capable d'effectuer la reconnaissance d'un modèle de visage 3D aussi dans les cas où seulement une partie du modèle est disponible. Les expériences de Berritti, Del Bimbo rapportés un taux de reconnaissance de rang 1 sur la base de données FRGC v1.0 atteint 70 %. Cependant, le réglage expérimental utilisé dans ces expériences suppose implicitement que les régions ne sont pas comparées entre eux ou la classification des données de chaque partie de visage est absente. Huang et al. [103] présentent une méthode de reconnaissance faciale 3D basée sur les différences de statistiques d'images de profondeur. Un TEE de 12,4 % est atteint dans l'expérience de ROC III sur la base de données FRGC v.2.

Wang et al. [104,105] emploient la SSDM (Signed Shape Difference Map) calculée entre deux visages 3D alignés comme une représentation pour la comparaison de la surface du visage.

Tang et al. [106] développent un algorithme de reconnaissance faciale 3D en utilisant un LBP (Local Binary Pattern) en variétés d'expression, extension de l'opérateur LBP, qui est largement utilisée dans l'analyse du visage 2D. Tout d'abord, pour décrire le visage humain avec plus de précision et réduire l'effet de sa distorsion locale, un système de division 3D du visage est proposé. Ensuite, pour chaque région du visage, l'histogramme statistique est utilisé pour résumer les détails du visage. Enfin, l'algorithme 3D de reconnaissance faciale proposé est testé sur les bases de données BJUT-3D et FRGC v2. Les auteurs ont obtenu des résultats prometteurs et ont conclu qu'il est possible d'appliquer la représentation de LBP sur la reconnaissance faciale 3D.

2.3.2.2 Méthodes basées sur les régions

La totalité de la surface du visage peut être segmentée relativement en parties rigides et non rigides en attribuant un poids plus faible sur les régions du visage non rigides pour réduire leurs

effets. Les régions du visage sélectionnées sont ensuite comparées à un certain espace de caractéristiques pour la reconnaissance faciale.

Alyuz et al. [107] proposent une approche régionale de l'enregistrement qui est robuste face aux expressions pour la reconnaissance faciale 3D.

Deux types de descripteurs de forme 3D : coordonnées des points et courbures de surface faciale sont considérées. L'algorithme du plus proche voisin est utilisé. Deux bases de données du visage 3D : la FRGCv2 et le sous-ensemble de base de données Bosphorus hors variation de pose et d'occlusion sont utilisés pour l'évaluation expérimentale.

2.3.2.3 Méthodes basées sur le calcul de courbe

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

Samir et al. [108] ont présenté une nouvelle approche de comparaison de surfaces faciales 3D pour la reconnaissance. Un certain nombre de prétraitements était introduit avant cette opération, telle que la conversion du modèle 3D en une image de profondeur, le bouchage des trous, l'élimination des pics par lissage gaussien et finalement la segmentation par un masque elliptique de la surface faciale.

Zhang et al. [109] ont fait usage de modèles de visage 3D et leurs images pour extraire les profils de symétrie centrale. Sur la base des valeurs de courbure des profils, trois points essentiels sont situés sur chaque surface du visage. En expérimentant avec un ensemble de modèles représentant différentes expressions faciales. Mahoor et al. [110] introduisent la courbure principale, k_{max} , pour représenter l'image du visage comme une image 3D binaire appelée image de crête. L'image de crête indique l'emplacement des lignes de crête autour des régions faciales importantes sur le visage (les yeux, le nez et la bouche).

Jahanbin et al. [111] décrivent les surfaces 3D par iso-profondeur et courbes iso-géodésique. La dernière étape consiste à vérifier ou rejeter les demandes des utilisateurs à l'aide d'un classificateur de distance euclidienne ou SVM.

Drira et al. [112] ont exploré les courbes radiales élastiques (Elastic Radial Curves) pour modéliser les déformations faciales 3D causées par les variations d'expression du visage. La distance géodésique entre les courbes est adoptée comme mesures de similarité pour comparer les surfaces faciales.

2.3.2.4 Méthodes basées sur le calcul du point

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

Wu et al. [113] proposent LSM (Local Shape Map) pour décrire un point sur la surface du visage. Cette méthode ne nécessite pas d'enregistrer deux surfaces faciales. Castellani et al. [114] introduisent une méthode d'apprentissage générative par adaptation de la technique HMM qui travaille sur des maillages 3D. Daniyal et al. [115] forment un modèle dit du point de distribution (Point Distribution Model, PDM). Toutes les distances euclidiennes entre tous les deux points sont exploitées en tant que caractéristiques et classées par la méthode LDA pour la reconnaissance 3D.

Maes et al. [116] étendent l'algorithme SIFT à ce qui est connu sous le nom Mesh SIFT. Contrairement aux informations du gradient extraites par l'opérateur SIFT original, Mesh SIFT code l'histogramme local de l'indice de forme. Les tests sont effectués sur l'ensemble des données de Bosphorus et atteignent un taux de reconnaissance égal à 93,7 %.

Berretti et al. [117] ont proposé un système de reconnaissance de visage 3D dans le cas où seulement des parties des scans de test sont disponibles. Des points-clés détectés sur les scans du visage de cinq personnes différentes pour la combinaison de la base de données FRGC v2 et UND.

2.3.2.5 Les méthodes basées sur la fusion des caractéristiques multiples

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

Gokberk et al. [118] comparent cinq approches de reconnaissance faciale 3D. Ils comparent les méthodes basées sur l'EGI, la correspondance par ICP, la profondeur de profil, la PCA et la LDA. Ils explorent également la fusion des résultats des cinq approches qui sont capables d'atteindre 99 % de taux de reconnaissance. À ce moment, ce travail est relativement nouveau dans la comparaison des performances des différents algorithmes de reconnaissance faciale 3D.

Moreno et al. [119] ont calculé trente caractéristiques géométriques locales contenant la distance, l'angle, l'aire, les courbures... etc. pour modéliser les surfaces faciales 3D. Les techniques PCA et SVM sont utilisées pour la classification sur l'ensemble de données Gavab. Lee et al. [120] ont détecté manuellement huit points, sept courbes ainsi que des valeurs de distance et d'angle.

Li et Zhang [121] ont exploré l'utilisation de plusieurs attributs géométriques intrinsèques, y compris les angles, les distances géodésiques et les courbures pour la reconnaissance faciale 3D, où chaque face est représentée par un maillage triangulaire, prétraité pour posséder une connectivité uniforme. En utilisant l'approche de visage propre basée sur les poids appris et un classifieur de voisin le plus proche. Les résultats sur les bases de données Gavab et FRGC montrent que la fusion améliore la performance.

Guan et Zhang [122] combinent FSA (Facial Structural Angle) et LRM (Local Region Map) pour la reconnaissance faciale 3D. Les auteurs affirment que cette stratégie est robuste par rapport aux expressions faciales.

2.3.3 Autres approches [4]

En plus des techniques globales et locales de reconnaissance du visage 3D, il existe aussi des approches visant à trouver une relation entre deux surfaces 3D du visage dans un espace de cartographie spéciale.

Compte tenu de ces approches, nous les classons comme méthodes de para métrisation de surface. Un para métrisation d'une surface peut être considéré comme une mise en correspondance "un contre un" de la surface dans un domaine approprié. Les surfaces faciales 3D qui sont représentées sous forme de maillages peuvent être mappées sur une image 2D régulière de sorte que des approches plus sophistiquées de traitement du signal peuvent être appliquées.

Pan et al. [123] explorent une profondeur d'image mappée pour la reconnaissance faciale 3D. Avec l'utilisation de la pointe du nez comme point central et un axe de symétrie pour l'alignement, les données de face sont mappées sur une plage de l'image circulaire. L'approche, basée sur la technique PCA, fut introduite pour la classification. La performance rapportée est 95 % de taux de reconnaissance et un TEE= 2,8 % dans un scénario de vérification sur l'ensemble de données FRGC v1.

Passalis et al. [124] adaptent d'une part un modèle déformable prédéfini annoté manuellement d'une surface donnée du visage 3D de façon à pouvoir être paramétrée sur un plan 2D pour produire une image en couleurs, dont trois canaux contiennent les coordonnées des points de l'ensemble des sommets.

Les expériences sur la base de données FRGC v2 montrent que cette méthode fonctionne beaucoup mieux que la technique PCA et qu'elle est également insensible aux variations d'expression et de genre.

2.3.4 Discussion

L'objectif actuel des techniques de reconnaissance faciale 3D est de répondre aux exigences de l'application en conditions réelles, c'est-à-dire de fournir des précisions élevées. Les variations d'expression de visage représentent le principal facteur qui dégrade largement la précision d'un système de reconnaissance 3D, pendant ce temps, l'étape de prétraitement joue un rôle important dans la reconnaissance du visage.

Comme les données 3D donnent des informations plus utiles que les données 2D, le prétraitement à l'aide des données 3D est plus pratique et précis. En règle générale, dans l'analyse de visage 3D, l'étape de prétraitement comporte deux parties principales, à savoir la détection des points de repère et l'enregistrement de la surface du visage.

Quelques points de repères clés sont essentiels pour estimer les positions initiales des surfaces du visage ou localiser les traits du visage les plus complexes (par exemple, les courbes et les régions, etc.).

2.4 Reconnaissance de l'expression faciale

2.4.1 Qu'est-ce qu'une expression faciale ? [131]

Le visage humain représente une ressource d'informations très riche dont la forme et le comportement de chaque membre changent son apparence. Dans la science anatomique, le visage comporte 44 muscles qui peuvent être activés volontairement ou non pour engendrer un nombre total égal à 250 000 expressions faciales [125].

Une expression faciale est définie dans le dictionnaire Hachette (édition 1997) comme étant une *"manifestation d'une pensée, d'un sentiment par le visage"*. À partir de ce qui précède, une expression faciale est alors un changement de position et de forme d'un ou de plusieurs traits du visage par rapport à la position et la forme par défaut. Ce changement est le résultat de l'activation des muscles relatifs aux traits faciaux selon ou non la volonté de la personne.

En 1975, P. Ekman et W. V. Friesen [126] ont montré la présence de sept expressions du visage qui sont la surprise, la peur, le dégoût, la colère, la joie, la tristesse et la neutralité. Les six premières expressions sont dites les six universelles et elles sont identifiables par n'importe quelle personne indépendamment de sa culture [127]. Les mêmes physiologistes ont inventé en 1978, un système de codification manuelle des expressions faciales, connu sous le nom « FACS : Facial Action Coding System ». Il représente tous les mouvements visibles du visage à l'aide d'une combinaison entre une ou plusieurs actions unitaires (Action Unit).

Une action unitaire décrit le mouvement élémentaire des muscles entrant un changement sur l'apparence de visage par exemple le relèvement des sourcils consiste en une position actuelle des sourcils plus haute que sa position précédente [128].

La figure 11 regroupe certains muscles de visages et leurs rôles et quelques exemples des actions unitaires. Essa et al. [129] ont proposé une version améliorée « FACS⁺ » du système FACS. Ce nouveau système facilite l'analyse et le codage d'une expression faciale par l'ordinateur.

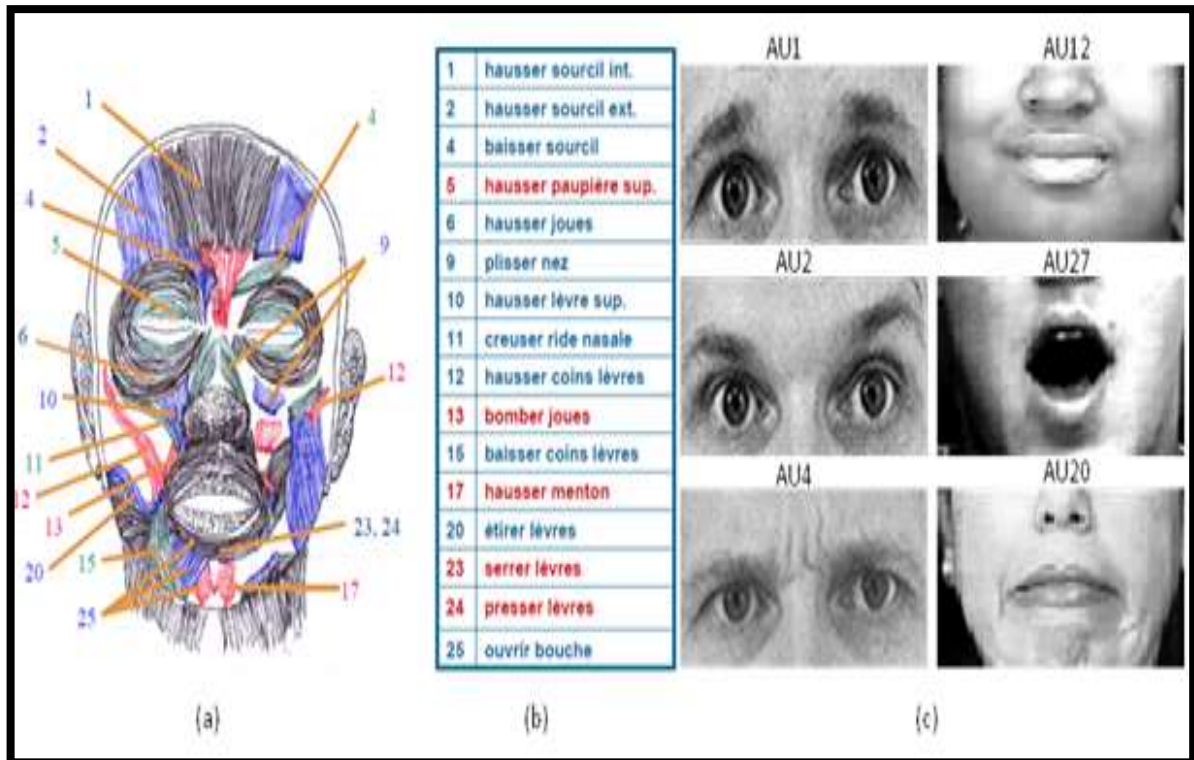


Figure 2.8. (a) Extrait de muscles de visage, (b) définitions des rôles des muscles d'après [125] et (c) exemples des actions unitaires du système FACS d'après [126] et [127].

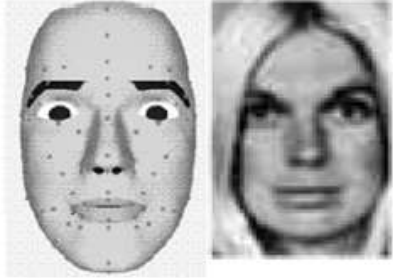
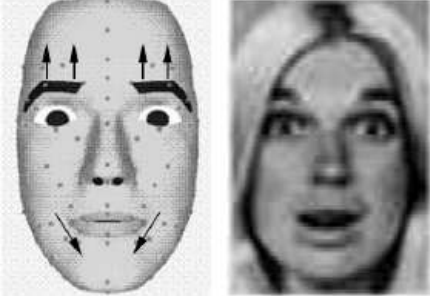
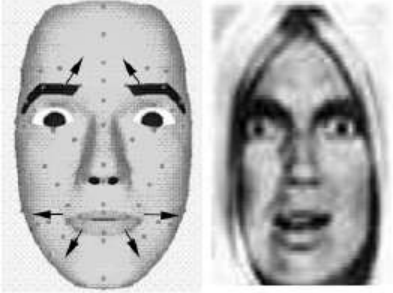

Afin de classer correctement une expression faciale requête, nous allons donner une brève description de chacune des expressions mises au point par Paul EKMAN afin de fixer leurs caractéristiques pertinentes.

Nous nous basons sur les définitions présentes dans le dictionnaire encyclopédique Hachette, sur les travaux de Khadoudja Ghanem [130] et sur les images présentées par Essa et *al.* [129, 125] exposant les mouvements faciaux globaux.

Les tableaux ci-dessus montrent la description des différentes expressions d'après [130], [125] et [129].

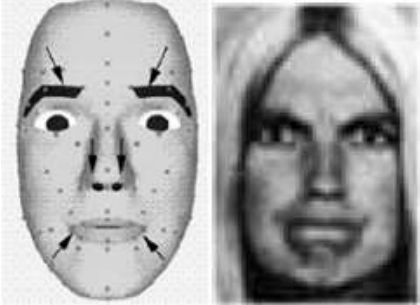


Chapitre 2 : La reconnaissance faciale et ses différentes techniques

Tableau 2.1. Description des sept expressions d'après [130], [125] et [129]. [131]

Expression \ Caractéristiques	Description de la classe d'expression
<p><i>Faciale</i></p> <p><i>Neutre</i></p> 	<ul style="list-style-type: none"> ❖ Définition : Etat d'une personne qui n'a pas d'expression marquée et n'a pas de changement sur son visage. C'est l'état de référence. ❖ Critères : Les différents traits de visage sont à leurs positions et formes par défaut.
<p><i>Surprise</i></p> 	<ul style="list-style-type: none"> ❖ Définition : Etat d'une personne étonnée par quelque chose d'inattendu. ❖ Critères : <ul style="list-style-type: none"> • Les sourcils sont élevés. • Les yeux sont grands ouverts. • La bouche est ouverte verticalement (la hauteur augmente et la largeur diminue).
<p><i>Peur</i></p> 	<ul style="list-style-type: none"> ❖ Définition : Etat d'une personne menacée par un danger réel ou imaginaire. ❖ Critères : <ul style="list-style-type: none"> • Les sourcils sont légèrement élevés. • Les yeux sont écartelés. • La bouche est ouverte horizontalement (hauteur et largeur augmentent moyennement).
<p><i>Dégout</i></p> 	<ul style="list-style-type: none"> ❖ Définition : Etat d'une personne qui a une répugnance pour certains aliments ou a un manque d'appétit. ❖ Critères : <ul style="list-style-type: none"> • Les coins intérieurs des sourcils sont légèrement descendus. • Les yeux sont semi-ouverts. • La bouche est fermée et déplacée vers le haut.

Chapitre 2 : La reconnaissance faciale et ses différentes techniques

Tableau2.2. Description des sept expressions d'après [130], [125] et [129]. [131]

<p><i>Colère</i></p> 	<ul style="list-style-type: none">❖ Définition : Etat d'une personne dans une réaction violente et agressive lors d'une contrariété.❖ Critères :<ul style="list-style-type: none">• Les sourcils sont plissés et ont la tendance à se rallier. Leurs parties intérieures sont peu descendues.• Les yeux sont quasiment fermés.• La bouche est fermée, serrée et à peine déplacée vers le haut.
<p><i>Joie</i></p> 	<ul style="list-style-type: none">❖ Définition : Etat d'une personne dans une condition de satisfaction intense.❖ Critères :<ul style="list-style-type: none">• Les sourcils sont faiblement étirés.• Les yeux sont quasiment plissés (la paupière inférieure cache une partie de l'œil).• La bouche est ouverte horizontalement et verticalement.
<p><i>Tristesse</i></p> 	<ul style="list-style-type: none">❖ Définition : Etat d'une personne moralement souffrante suite à une insatisfaction et des soucis.❖ Critères :<ul style="list-style-type: none">• Les coins intérieurs des sourcils sont légèrement surélevés.• Les yeux sont semi-ouverts (les paupières cachent une partie du champ de vision).• La bouche est serrée et peu ouverte (les coins de la bouche sont étirés vers le bas).

2.5 Reconnaissance de visage 2D+3D [4]

Cherchant une solution aux problèmes qui affectent de manière significative la reconnaissance faciale 2D, en particulier la sensibilité aux changements de pose, d'éclairage, les chercheurs ont utilisé les modèles de visage 3D depuis 1989 et y ont contribué par des améliorations significatives. Fort heureusement, la plupart des scanners de visage 3D (par exemple les scanners stéréoscopiques) capturent simultanément des images 2D sous des angles différents et les utilisent comme des données brutes dans leur méthode de calcul de la profondeur. Il s'agit de techniques qui combinent des données 2D et 3D sur le visage pour améliorer les performances et la robustesse de la reconnaissance. Bien que la reconnaissance du visage 3D soit ancienne d'environ deux décennies, la reconnaissance multimodale de visage 2D+3D est au stade préliminaire et a eu une exposition limitée dans la littérature depuis l'année 2000. Récemment, plusieurs approches basées sur ce principe ont été développées, la majorité des algorithmes de cette catégorie utilisent des méthodes classiques, comme les visages propres, pour créer des systèmes de reconnaissance indépendants 2D et 3D et pour combiner les scores de concordance par une stratégie de fusion.

Bronstein et al. [132] ont proposé une méthode qui considère le visage comme une surface isométrique. Ils se sont inspirés des travaux de Schwart et al. [08], les premiers à employer la graduation multidimensionnelle (MDS) pour projeter des surfaces corticales du cerveau sur un plan. Bronstein et al. [134] proposent un schéma de décomposition utilisant la technique PCA multimodale 2D+3D. Beumier et al. [135] proposèrent un nouveau schéma de reconnaissance multimodale par l'introduction d'une comparaison de profils 2D combinée avec la première approche. La fusion est effectuée moyennant une somme pondérée de scores élémentaires.

Chang et al. [136] ont mené une étude expérimentale poussée qui leur a permis de comparer et de combiner des données 2D et 3D pour la reconnaissance faciale. Les sujets ont été scannés à l'aide d'un scanner laser "Minolta Vivid 900" qui permet d'obtenir des images d'intensité 2D et de profondeur 3D. Dans leur expérimentation, les auteurs ont utilisé la distance de Mahalanobis pour réaliser la mise en correspondance (Matching). Wang et al. (2002) [137], ont travaillé sur les descripteurs locaux des deux images de profondeur et d'intensité sont intégrés au niveau de données. Les vecteurs caractéristiques 2D et 3D sont initialement projetés à l'aide de la PCA. Deux méthodes de classification sont évaluées : une mesure de similarité et une SVM multi-class.

Cook et al. [138] ont proposé un algorithme pour combiner les données de visage 2D et 3D pour améliorer la précision de reconnaissance. Dans cet algorithme, les images de profondeur et d'intensité sont convoluées avec les filtres log-Gabor. Les distances de Mahalanobis résultant de la comparaison sont recueillies dans un vecteur de caractéristiques utilisé par une machine à vecteurs de support linéaire (SVM).

Mpiperis et al. [139] proposèrent un algorithme de reconnaissance de visage 2D+3D où les expressions du visage (images de profondeur et d'intensité) sont compensées par le paramétrage polaire géodésique de la surface du visage.

Al-Osaimi et al. [140] proposent la fusion optimisée au niveau de données (pixel-level fusion) 2D et 3D pour la reconnaissance faciale. Les fonctions de fusion sont objectivement optimisées pour modéliser les variations d'expression et d'éclairage de sous-espaces linéaires invariants de

Chapitre 2 : La reconnaissance faciale et ses différentes techniques

reconnaissance de visage. Les expériences sur la base de données FRGC v2 montrent que l'optimisation spatiale et la combinaison de plusieurs de ces fonctions de fusion améliorent systématiquement les performances.

Wang et al. [141] utilisent la méthode "Adaboost" pour l'extraction de la région faciale. Puis l'extraction de caractéristiques 2D et 3D du visage est faite par DTCWT (Dual-tree Complex Wavelet Transform). En conclusion, la plupart des approches multimodales (2D+3D) confirment l'intérêt de la fusion des deux modalités en vue d'obtenir des taux de reconnaissances et d'authentification plus élevés que les approches 2D ou 3D seules.

2.6 Conclusion

Dans la première partie de ce chapitre, nous avons examiné plusieurs techniques utilisées pour la reconnaissance de visage 2D en fonction de l'image d'intensité. Nous avons également examiné, dans la seconde partie, les approches de représentation 3D basées exploitées dans la reconnaissance faciale 3D. Enfin, dans la troisième partie nous avons discuté les modèles de reconnaissance faciale multimodale (2D+3D).

Avec la modalité 2D, presque toutes les méthodes récentes affirment être en mesure de fournir des taux de reconnaissance satisfaisants, mais seulement quand elles sont évaluées sur des bases de données standard alors que l'évaluation en temps réel. En fait, certains des défis fondamentaux, à savoir les variations d'éclairage et de pose, n'ont pas été complètement résolus.

La modalité 3D a la capacité d'améliorer les performances de reconnaissance faciale 2D par compte tenu de ces défis. De nombreuses techniques ont déjà prouvé leur capacité dans certains cas de validation. Cependant, plusieurs autres problèmes difficiles ont encore besoin d'être dûment pris en compte, comme la précision, la vitesse d'acquisition des données, la précision de détection des points clés et la lenteur d'enregistrement de la surface. Enfin, la modalité 3D a tendance à présenter plus de sensibilité aux variations des expressions du visage par rapport à celles basées sur des images 2D. Pour améliorer la performance dans chaque modalité, plusieurs chercheurs ont proposé de meilleures solutions.

Nous avons clôturé ce chapitre par un résumé des méthodes de fusion (2D et 3D) afin que la robustesse des systèmes de reconnaissance faciale puisse être améliorée.

Notre système pour la reconnaissance de visage en 2D

3.1 Introduction

Durant les vingt dernières années, des efforts importants ont été consentis dans le domaine de la reconnaissance automatique de visages, et plusieurs techniques ont été développées. Ces techniques, détaillées dans le chapitre précédent, sont regroupées, selon quelques critères, en deux catégories principales : les techniques dites globales et les techniques basées sur des caractéristiques locales.

La reconnaissance faciale par ordinateur, tâche que les humains effectuent naturellement et sans effort dans leur vie quotidienne consiste à transposer cette caractéristique de reconnaissance à une machine à l'effet de la réaliser d'une manière systématique et automatique par un dispositif de traitement matériel et logiciel.

La grande disponibilité d'ordinateurs puissants ainsi que des systèmes informatiques embarqués ont suscité un énorme intérêt dans le traitement automatique des images au sein de nombreuses applications, incluant l'identification biométrique, la surveillance automatique, l'interaction homme-machine et la gestion de données multimédia. Nous allons donc présenter dans ce chapitre notre système monomodal et multi-algorithmes de vérification de visage 2D.

Dans la partie de vérification de visage 2D monomodale ou uni-modale, nous présentons tout d'abord notre modèle de vérification de visage, puis nous détaillons les quatre étapes principales de ce modèle. Enfin, nous étudions l'extraction des caractéristiques puis nous détaillons l'étude expérimentale sur le système automatique 2D. Celle-ci passe par la comparaison des algorithmes de réduction de l'espace de données, par la technique LBP (Local Binary Patterns) plus MB-LBP (Multi Block- Local Binary Patterns). Enfin, l'étude de l'algorithme SVM pour l'apprentissage intégrée dans la bibliothèque OpenCV.

Dans la partie de vérification de visage 2D multi-algorithmes, les limitations des systèmes biométriques lorsqu'ils n'utilisent qu'une seule modalité biométrique sont passées en revue. Nous présentons ensuite les différentes multimodalités possibles. Nous terminons par une description de la reconnaissance d'expressions faciales et nous clôturons notre chapitre par une conclusion.

3.2 Vérification uni-modale de visage 2D

Le principe de fonctionnement de base de notre système de reconnaissance de visage est d'exploiter un flux composé simultanément d'une image couleur capturée par une caméra. Ce système de vérification faciale (figure 3.1) peut être résumé en quatre étapes : prétraitement, extraction de caractéristiques, réduction d'espace et comparaison.

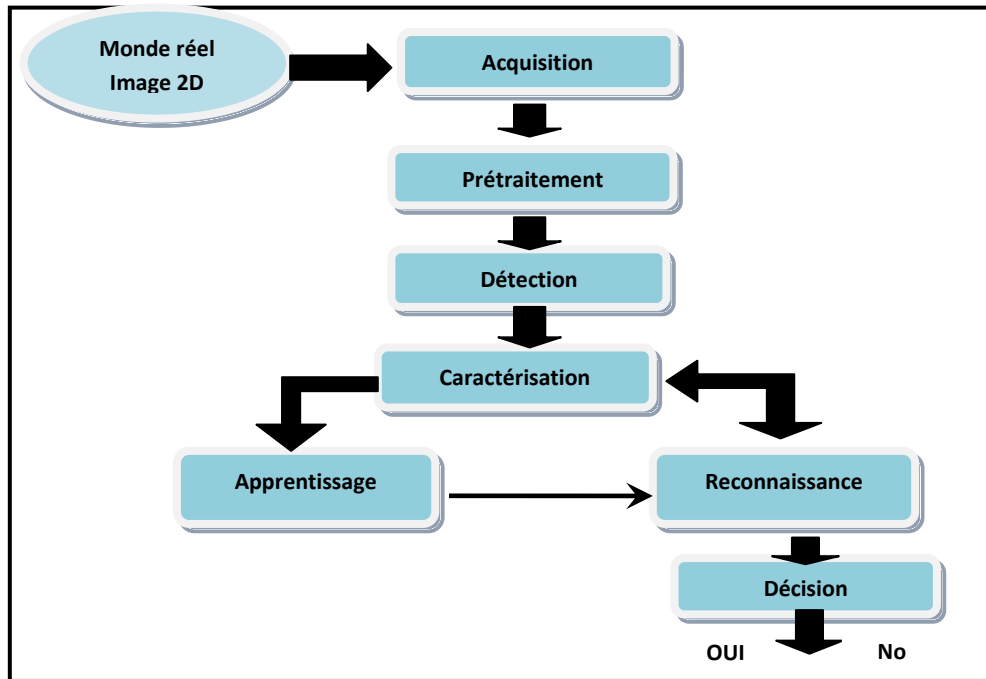


Figure 3.1. Schéma général de notre approche.

Dans ce qui suit, nous allons détailler chacune de ces étapes.

3.2.1 Acquisition de données

C'est la première étape de notre système qui consiste à acquérir des données à partir du monde réel. Dans notre cas, toutes les données d'entrées sont acquises via une caméra, celle de l'ordinateur, par exemple. Ce dernier tolère des captures simultanées d'une image couleur 2D.

Un système d'acquisition équipé d'un capteur est utilisé pour acquérir une caractéristique spécifique de l'utilisateur, par exemple : un microphone dans le cas de la voix. C'est l'opération qui permet d'extraire une représentation bidimensionnelle du monde réel pour des objets en 3D. Cette opération peut être statique (Appareil photo, Scanner, etc.) ou dynamique (Caméra, Web Cam), dans ce cas on aura une séquence vidéo. L'image acquise à ce niveau est une image brute.

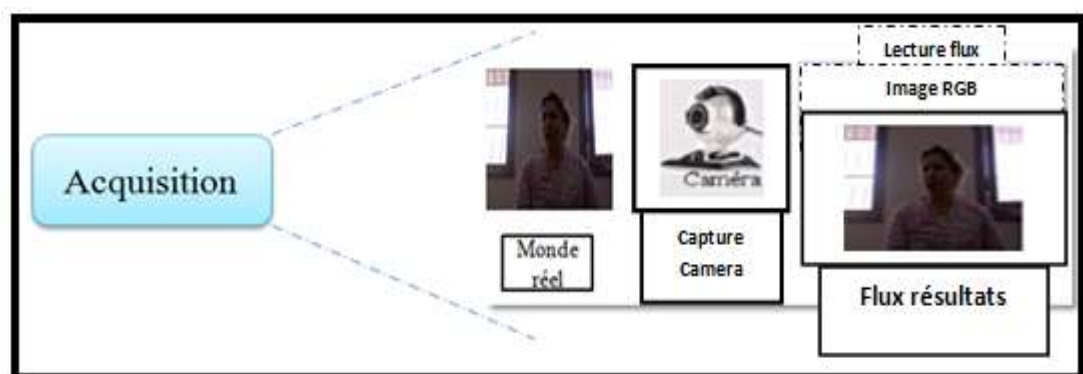


Figure 2 : Détails de l'étape « Acquisition ».

3.2.2. Prétraitement d'image

3.2.2.1 Prétraitement

Le prétraitement est une phase importante dans le processus global d'authentification. C'est une méthode simple qui augmente en général les performances du système. Elle permet souvent une première réduction des données et atténue les effets d'une différence de conditions lors des prises de vues.

Dans notre travail, nous utilisons trois phases de prétraitement : le découpage de l'image, la décimation et la conversion de la couleur dans l'espace de couleur TSV [4].

3.2.2.2. Découpage

Le découpage de l'image consiste à conserver les maximums des variations intrinsèques du visage, et à supprimer les autres informations comme l'arrière-plan, les cheveux, les cols de chemise, les oreilles...etc. afin d'améliorer les performances du système d'authentification. Une fenêtre rectangulaire de taille (161 × 121), centrée autour des caractéristiques les plus stables liées aux yeux, aux sourcils, au nez et à la bouche, a été utilisée. La figure 3.3 montre l'utilisation de cette fenêtre [4].



Figure 3.3. Image de visage (a) avant et (b) après découpage.

3.2.2.3 Décimation

La décimation consiste à ne prendre qu'un pixel sur quatre. Ceci réduit bien entendu la résolution des images. L'image de visage passera ainsi d'une dimension $256 \times 256 = 65536$ pixels à une dimension de $81 \times 61 = 4941$ pixels après découpage et décimation, comme illustré par la figure 3.4.



Figure 3.4. Image de visage (a) avant (b) après décimation.

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

Les systèmes d'authentification de visages utilisent souvent la représentation en niveaux de gris de l'image du visage comme caractéristique d'entrée. Cependant, lorsque les images sont en couleur, il est avantageux d'utiliser cette information supplémentaire pour améliorer les performances de ce système d'authentification. Nous utilisons donc l'espace de couleur TSV (Teinte, Saturation, Valeur) ou HSV en anglais (Hue, Saturation, Value). Ce choix est justifié dans [142] où l'on confirme que c'est le système le plus utilisé dans la littérature. La quantification de la couleur de notre base de données est faite dans le système RVB (Rouge, Vert et Bleu). Le passage du système de couleurs RVB au système TSV se fait par une fonction de transfert non linéaire. La figure 3.4 présente les deux systèmes colorimétriques RVB et TSV [4].

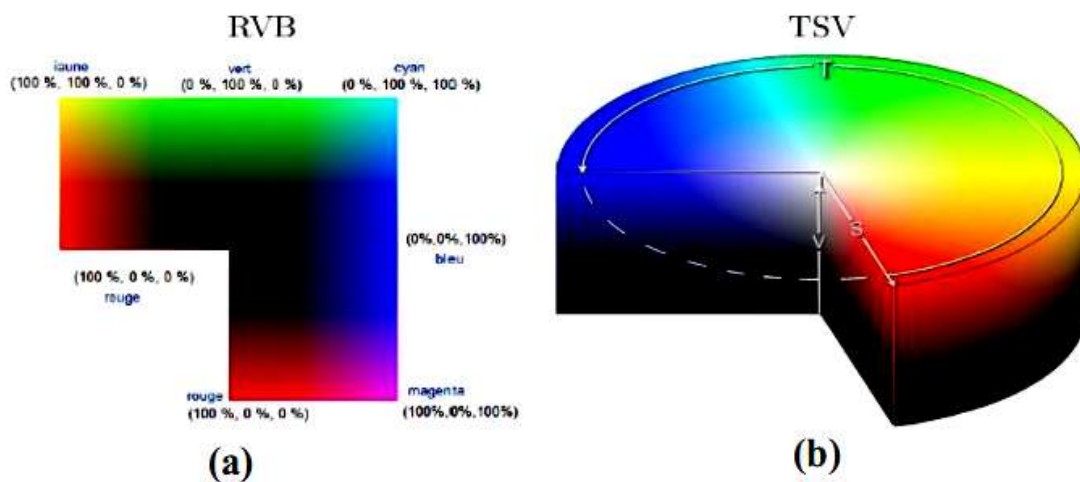


Figure 3.5. Différents systèmes colorimétriques : (a) plan du système de couleurs RVB, (b) cylindre du système de couleurs TSV.

Soient min et max, respectivement, les valeurs minimale et maximale des composantes R, V et B (min = minimum (R, V, B) et max = maximum (R, V, B)) et soient x et y les coordonnées d'un pixel de l'image. Les transformations des espaces RVB vers TSV sont données par les équations suivantes [142]. (Figure3.6)

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

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

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

Figure 3.6. Les équations des transformations des espaces RVB vers TSV.

Pour plus de détails sur cette transformation, on peut se référer à [143]. Dans [136], on a étudié les espaces de couleurs sur la base de données XM2VTS et on a trouvé que la composante S donne

le meilleur résultat pour la tâche de vérification d'identité des visages. Partant de ce constat, nous utilisons donc la composante S de l'espace de couleur TSV. La figure 3.7 présente la conversion de couleur et l'extraction de la composante S du visage [4].

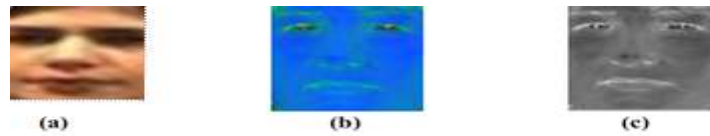


Figure 3.7. (a) Image découpée et décimée, (b) Image en système TSV, (c) Composante S.

3.2.3 Détection

La détection automatique du visage est la première et l'essentielle étape pour toute application de reconnaissance des visages. Elle est composée de deux tâches : au cours de la première tâche, l'application indique si l'image contient un visage ou non et au cours de la seconde, l'application isole la région contenant le visage s'il est présent.

En général, les systèmes de détection utilisent les sous-fenêtres lors de la recherche du visage dans l'image. Cette recherche doit être effectuée de manière exhaustive avec le minimum de calcul ainsi qu'avec le minimum de temps. Le résultat de cette recherche est une liste de zones susceptibles de contenir un visage. Ces zones sont recherchées encore une fois par des méthodes plus robustes.

3.2.4 Caractérisation

L'étape de caractérisation consiste à définir les caractéristiques pertinentes de et d'éliminer les informations redondantes. Puisque nous faisons usage de la méthode « eigenface », la première tâche de cette étape consiste à générer un ensemble de visages propres dont le nombre est égal au nombre d'images de la base. La deuxième tâche est le maintien des X premiers visages propres générés qui disposent de l'information caractéristique de visage. La dernière tâche est le calcul de la matrice des poids dans l'espace engendré par les visages propres retenus.

Pour la réduction d'espace engendré des images d'apprentissage nous faisons l'usage de la méthode « LBP » ensuite la méthode « MB_LBP ». Parce que la LBP met en évidence les caractéristiques locales lisses de l'image et l'intensité du visage, et pour plus d'optimisation nous avons utilisé MB_LBP.

❖ La réduction d'espace

Dans ce chapitre, nous avons utilisé des approches locales basées sur la réduction de l'espace de données, car elles restent parmi les techniques les plus utilisées et les plus populaires dans le domaine de la reconnaissance des visages. Dans ces procédés, les images du visage sont traitées de manière locale et sont généralement transformées en vecteurs, qui sont plus faciles à manipuler par les méthodes suivantes :

- Local Binary Patterns (LBP) : Récemment, la représentation LBP a gagné beaucoup d'attention dans le domaine de la reconnaissance faciale et est également appliquée avec

succès en reconnaissance faciale 2D. Inspirés par l'opérateur LBP 2D, nous avons utilisé un algorithme de reconnaissance de visage 2D.

- Multi block LBP, (MB-LBP), qui, au lieu de comparer les pixels, compare les intensités moyennes des sous-régions voisines.

Le but principal de ces méthodes est de trouver une base de projection qui optimise un certain critère. Cette base est ensuite projetée sur une base de dimension inférieure. Dans notre travail, nous avons testé ces méthodes sur notre base de données générée du visage 2D. Nous avons montré que la méthode MB_LBP est meilleure que d'autres méthodes de réduction d'espace.

❖ Technique LBP [4]

L'opérateur LBP a été proposé initialement par Ojala et al. [146] dans le but de caractériser la texture d'une image. Le calcul de la valeur LBP consiste pour chaque pixel à "seuiller" ses huit voisins avec un seuil dont la valeur est le niveau de gris du pixel courant. Tous les voisins prendront alors une valeur 1 si leur valeur est supérieure ou égale au pixel courant et 0 si leur valeur est inférieure (figure 3.8). Le code LBP du pixel courant est alors produit en concaténant ces 8 valeurs pour former un code binaire. On obtient donc, comme pour une image à niveaux de gris, une image des valeurs LBP contenant des pixels dont l'intensité se situe entre 0 et 255.

La technique LBP a été étendue ultérieurement en utilisant des voisinages de tailles différentes. Dans ce cas, un cercle de rayon R autour du pixel central est considéré. Les valeurs des P points échantillonnés sur le bord de ce cercle sont prises et comparées avec la valeur du pixel central. Pour obtenir les valeurs des P points échantillonnés dans le voisinage pour tout rayon R, une interpolation est nécessaire. On adopte la notation (P, R) pour définir le voisinage de P points de rayon R d'un pixel. La figure 3.9 (a) illustre trois voisinages pour des valeurs de R et P différentes.

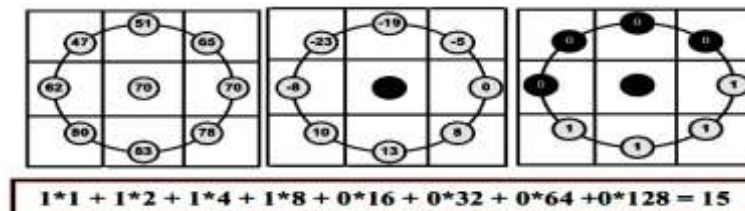


Figure 3.8. L'opérateur LBP

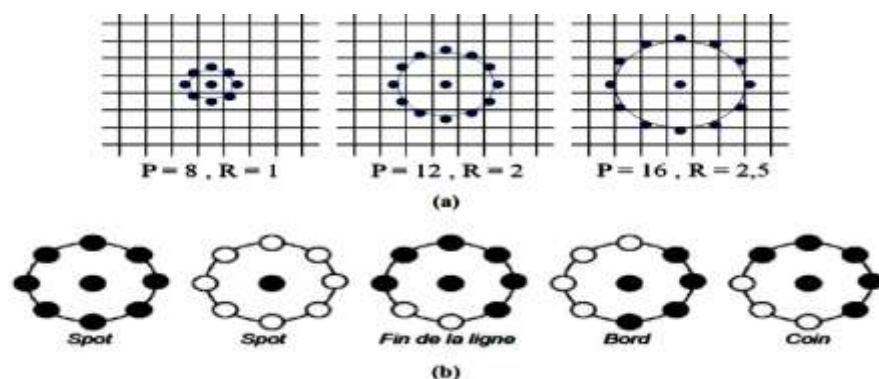


Figure 3.9. (a) Trois voisinages pour des R et P différents, (b) Textures particulières détectées par LBP.

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

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

$$\text{Où : } s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} \quad (1)$$

Où (x_c, y_c) sont les coordonnées du pixel courant, $LBP_{P,R}$ est le code LBP pour le rayon R et le nombre de voisins P . L'opérateur LBP obtenu avec $P = 8$ et $R = 1$ ($LBP_{8,1}$) est très proche de l'opérateur LBP d'origine. La principale différence est que les pixels doivent d'abord être interpolés pour obtenir les valeurs des points sur le cercle (voisinage circulaire au lieu de rectangulaire).

La propriété importante du code LBP est que ce code est invariant aux changements uniformes globaux d'illumination parce que la LBP d'un pixel ne dépend que des différences entre son niveau de gris et celui de ses voisins. Nous utilisons la LBP uniforme avec un nombre de points égal à 8.

➤ *LBP uniforme*

Un sous-ensemble de $2P$ codes binaires, appelé modèles uniformes et définis dans [147], peut être utilisé pour représenter les spots, les fins de ligne, les bords et les coins. La mesure de l'uniformité, $U(x)$, présentée par l'équation 4, enregistre le nombre de transitions spatiales dans la configuration binaire et le modèle uniforme qui contient au plus deux transitions de bits, c'est-à-dire $U(x) \leq 2$. Le modèle uniforme contient au total $(P - 1)P + 2$ codes binaires. L'opérateur LBP uniforme ($LBP_{P,R}^{u2}$) est défini comme suit :

$$(LBP_{P,R}^{u2})^n(x, y) = \begin{cases} I(LBP_{P,R}(x, y)) & \text{si } U(LBP_{P,R}(x, y)) \leq 2, I(z) \in [0, (P - 1)P + 2] \\ (P - 1)P + 2 & \text{si non} \end{cases} \quad (2)$$

$$\text{Où : } U(LBP_{P,R}) = |s(g_{p-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^P |s(g_p - g_c) - s(g_{p-1} - g_c)|$$

Où u_2 représenté dans l'expression 5.7 indique que la définition se rapporte à des modes uniformes avec une valeur $U \leq 2$. Si $U(x)$ est inférieur à 2, le pixel en cours est marqué par une fonction d'indexation $I(z)$. Sinon, il sera assigné à la valeur $(P - 1)P + 2$.

La fonction d'index, $I(z)$, contenant $(P - 1)P + 2$ indices, est utilisée pour attribuer un index particulier à chacune des configurations uniformes. Un exemple de codes uniformes pour $P=8$ est présent dans la figure 3.10.

L'utilisation d'un code LBP uniforme a deux avantages. Le premier est le gain en mémoire et en temps de calcul. Le deuxième est que $LBP_{P,R}^{u2}$ permet de détecter uniquement les textures locales importantes, comme les spots, les fins de ligne, les bords et les coins (figure 3.9.b) pour des exemples de ces textures particulières. En effet, Ojala et al. ont montré que les LBPs uniformes contiennent plus de 90 % de l'information d'une image.



Figure 3.10. Un exemple de codes uniformes pour $P=8$.

❖ Multi block LBP, (MB-LBP)

Li et al. [148], ont proposé un LBP multibloc (multiblock LBP, MB-LBP), qui, au lieu de comparer les pixels, compare les intensités moyennes des sous-régions voisines. La technique LBP originale peut être considérée comme un cas particulier de la MB-LBP.



Figure 3.11. Exemple de la méthode MB_LBP pour un visage.

3.2.5 Apprentissage

Comme tout système de reconnaissance de données biométriques, notre système nécessite une phase d'apprentissage, en utilisant la méthode SVM pour les raisons suivantes : Rapide en temps de calcul, offre une bonne gestion de capacité mémoire, offre un taux de reconnaissance élevé,...

- Les machines à vecteurs de support (SVM)

Les machines à vecteurs de support (Support Vector Machine notées SVM), également appelées Séparateurs à Vaste Marges, introduites par Vapnik [149] sont utilisées dans la fusion de données dans les références [150, 151, 152, 153, 154, 155]. Ils ont pour but de définir un hyperplan de séparation entre les deux classes, hyperplan qui minimise l'erreur de classification sur un ensemble d'apprentissage [4].

L'idée principale est de projeter des données appartenant à des classes différentes, non linéairement séparables, avec l'espace d'entrée, dans un plus grand espace appelé espace de caractéristiques, de sorte que les données deviennent linéairement séparables. Dans cet espace, l'hyperplan optimal construit par la technique est utilisé pour calculer la fonction de classement séparant les classes.

Cas de deux classes : $\{(x_1, y_1), \dots, (x_n, y_n)\}$ où $y_i \in \{-1, +1\}$ et x_i entrée i (3)

Cette méthode d'apprentissage supervisée peut engendrer une séparatrice plus ou moins complexe selon la nature du noyau choisi. Le noyau le plus simple est le noyau linéaire qui consiste à chercher un séparateur linéaire dans l'espace à N dimensions des scores. Le but des fonctions noyau est de transformer l'espace initial (des scores à N dimensions) en un espace de dimension

plus grande dans lequel les données pourraient être linéairement séparables. La séparatrice est donc toujours linéaire dans l'espace transformé par la fonction noyau, mais ne l'est plus dans l'espace des scores.

Le but de la technique dite SVM est donc de trouver une séparatrice qui minimise l'erreur de classification sur l'ensemble d'apprentissage mais qui sera également performante en général sur des données non utilisées en apprentissage. Pour cela, le concept utilisé est celui de marge (d'où le nom de séparateurs à vaste marge). La marge est la distance quadratique moyenne entre la séparatrice et les éléments d'apprentissage les plus proches de celle-ci appelés vecteurs de support (figure 3.12). Ces éléments sont appelés vecteurs de support car ce sont uniquement sur ces éléments de l'ensemble d'apprentissage que la séparatrice est optimisée [156].

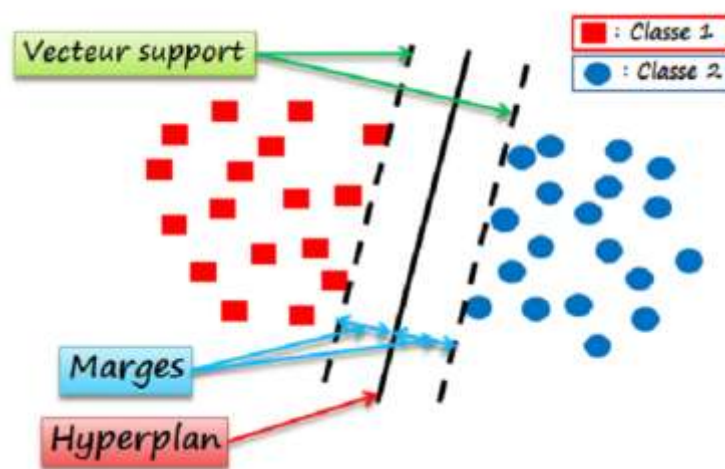


Figure 3.12. Séparation linéaire dans un espace à deux dimensions.

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

$$y = f(x) \quad (4)$$

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

$$w \cdot K(x) + b = 0 \quad (5)$$

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

$$h(x_i) = \frac{w \cdot K(x_i) + b}{\|w\|} \quad (6)$$

L'algorithme SVM est intégré dans la bibliothèque OpenCV.

3.2.6 Reconnaissance

C'est une comparaison du vecteur de poids test par rapport aux informations sauvegardées dans la matrice des poids issues de l'apprentissage L'objectif de cette étape est la détermination d'un visage existe bien dans notre référence. Pour le cas de notre approche, nous avons recourt à la distance euclidienne comme étant comparateur entre le vecteur requête et la référence. Avoir une distance minimale entre le vecteur test et une parmi les colonnes de la matrice de poids indique que l'image correspondante à ce dernier vecteur est la plus proche de l'image test.

Lorsqu'on souhaite comparer deux vecteurs de caractéristiques réduits issus du module de réduction de données d'un système biométrique, on peut, soit effectuer une mesure de similarité (ressemblance), soit une mesure de distance (divergence). Bien que la distance euclidienne soit optimale en théorie, diverses expériences ont constaté que celle-ci est surpassée par d'autres distances. Une d'entre elles est le cosinus [157] qui est définie par :

$$S(A, B) = \frac{A^T B}{\|A\| \|B\|} \quad (7)$$

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

3.3 Vérification multi-algorithmes de visage 2D [4]

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

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

3.3.1 Limitations des systèmes biométriques monomodaux

Bien que les techniques de reconnaissance biométrique promettent d'être très performantes, on ne peut garantir actuellement un excellent taux de reconnaissance avec des systèmes biométriques unimodaux. De plus, ces systèmes sont souvent affectés par les problèmes suivants [158] :

- Bruit introduit par le capteur : le bruit peut être présent dans les données biométriques acquises, ceci étant principalement dû à un capteur défaillant ou mal entretenu. Par exemple, un mauvais focus de caméra et le taux de reconnaissance d'un système biométrique sera très sensible.
- Non-universalité : si chaque individu d'une population ciblée est capable de présenter une modalité biométrique pour un système donné, alors cette modalité est dite universelle. Cependant, toutes les modalités biométriques ne sont pas vraiment universelles.
- Manque d'individualité : les caractéristiques extraites à partir de données biométriques d'individus différents peuvent être relativement similaires. Par exemple, (père et fils, vrais jumeaux, etc...). Ce manque d'unicité augmente le taux de fausses acceptations (TFA) d'un système biométrique.
- Manque de représentation invariante : les données biométriques acquises à partir d'un utilisateur lors de la phase de reconnaissance ne sont pas identiques aux données qui ont été utilisées pour générer le modèle de ce même utilisateur lors de la phase d'inscription. Ceci est connu sous le nom de variations intra-classe. Ces variations dues à une mauvaise interaction de l'utilisateur avec le capteur (par exemple, changements de pose et d'expression faciale lorsque l'utilisateur se tient devant une caméra), des algorithmes complexes sont requis pour prendre en compte ces variations. De grandes variations intra-classe augmentent le taux de faux rejet.
- Sensibilité aux attaques : Des études [159] ont montré qu'il était possible de fabriquer de fausses empreintes digitales en gomme et de les utiliser pour contrer un système biométrique. Les modalités biométriques comportementales telles que la signature et la voix sont plus sensibles à ce genre d'attaque que les modalités biométriques physiologiques.

3.3.2 Les différentes multimodalités possibles [4]

Les systèmes biométriques multimodaux diminuent les contraintes des systèmes biométriques monomodaux en combinant plusieurs systèmes. On peut différencier 5 types de systèmes multimodaux selon les systèmes qu'ils combinent [150], (figure 3.13) :

- multi-algorithmes : lorsque plusieurs algorithmes traitent la même image acquise, cette multiplicité des algorithmes peut intervenir dans le module d'extraction en considérant plusieurs ensembles de caractéristiques et/ou dans le module de comparaison en utilisant plusieurs algorithmes de comparaison.
- multi-instances : lorsqu'ils associent plusieurs instances de la même biométrie, par exemple l'acquisition de plusieurs images de visage avec des changements de pose, d'expression ou d'illumination.
- multi-capteurs : lorsqu'ils associent plusieurs capteurs pour acquérir la même modalité, par exemple un capteur optique et un capteur capacitif pour l'acquisition de l'empreinte digitale.
- multi-échantillons : lorsqu'ils associent plusieurs échantillons différents de la même modalité, par exemple deux empreintes digitales de doigts différents ou les deux iris. Dans ce cas les données

sont traitées par le même algorithme mais nécessitent des références différentes à l'enregistrement contrairement aux systèmes multi-instances qui ne nécessitent qu'une seule référence.

- multi-biométries : lorsque l'on considère plusieurs biométries différentes, par exemple visage et iris.

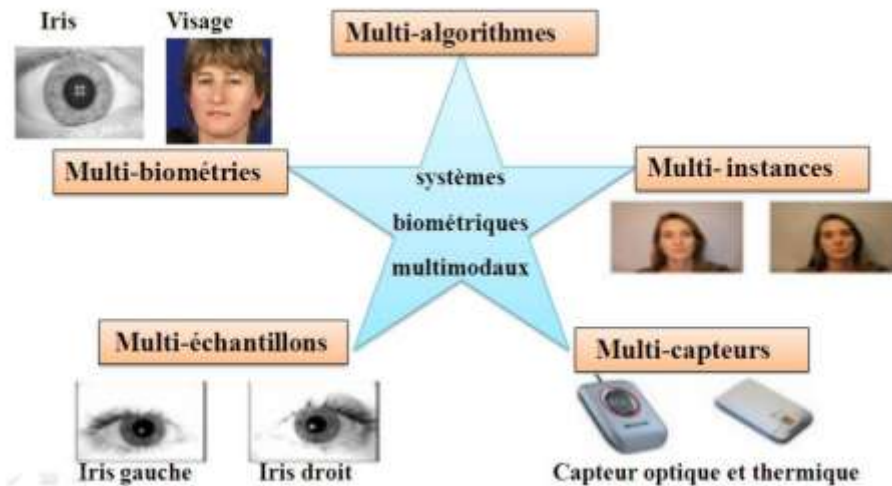


Figure 3.13. Les différents systèmes multimodaux [4].

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

3.3.3 État de l'art de la fusion de scores [4]

Au cours des dix dernières années, les systèmes multi-algorithmes ou multi-classifieurs (multiple classifier systems MCS) sont passés au statut de méthode établie pour concevoir des systèmes de classification. Un grand nombre d'éléments de preuves à la fois théoriques et expérimentales montrent que les systèmes multi-classifieurs peuvent surpasser un classifieur unique dans plusieurs applications réelles, en termes de précision de la classification (voir, par exemple [160, 161]).

En particulier, plusieurs chercheurs ont montré que la techniques MCS peut améliorer l'authentification biométrique de visages [162, 163, 164].

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

3.3.3.1 Approche par classification de scores

Plusieurs classifieurs ont été utilisés pour combiner les scores de correspondance afin d'arriver à une décision. Wang et al [165] considèrent les scores provenant de modules de reconnaissance faciale et de reconnaissance par l'iris comme un vecteur de caractéristiques à deux dimensions.

3.3.3.2 Approche par combinaison de scores

Kittler et al [166] ont développé un cadre théorique pour combiner les informations d'identification obtenues à partir de plusieurs classifieurs en utilisant des schémas tels que la règle « somme », la règle « produit », la règle « maximum », la règle « minimum » et la règle « médiane ».

Afin d'employer ces schémas, les scores doivent être convertis en probabilités à postériori conformément à un utilisateur client ou imposteur.

Plus récemment, Morizet et Allano ont essayé d'utiliser les deux approches et arrivèrent à des taux de réussite très intéressants.

3.3.4 La fusion de scores [4]

Il existe deux approches pour combiner les scores obtenus par différents systèmes : la première est de traiter le sujet comme un problème de combinaison, tandis que la deuxième est de voir cela comme un problème de classification. Dans l'approche par combinaison, les scores individuels sont combinés de manière à former un unique score qui est ensuite utilisé pour prendre la décision finale.

Afin de s'assurer que la combinaison de scores provenant de différents systèmes est cohérente, les scores doivent d'abord être transformés dans un domaine commun : on parle alors de normalisation de scores [167].

3.3.4.1 Normalisation de scores

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

Les méthodes de normalisation, présentées dans la suite, traitent des scores qui varient tous dans le même sens, on considère en général tous les scores sous forme de similarité [168]). La technique de normalisation de scores utilisés est :

- Normalisation par la méthode Min-Max [168, 34, 169]

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

$$S'_{ik} = \frac{s_{ik} - \min(\{s_i\})}{\max(\{s_i\}) - \min(\{s_i\})} \quad (8)$$

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

3.3.4.2 Approche par combinaison de scores

- Méthode de combinaisons simples

Les méthodes de combinaisons de scores simples sont des méthodes très simples dont l'objectif est d'obtenir un score final s à partir des N scores disponibles s_i pour $i = 1$ à N issus de N systèmes.

- Combinaison de scores par logique floue

La théorie de la logique floue (des sous-ensembles flous) a été introduite par Zadeh en 1965 [170] comme une extension de la logique binaire d'une part et une amélioration de la logique multi évaluée (multi-valued logics en anglais) d'autre part. L'importance de la logique floue réside dans le fait qu'elle s'approche du raisonnement humain par l'intégration et le traitement du caractère approximatif, vague, imprécis ou flou de la connaissance humaine.

3.3.4.3 Approche par classification de scores

L'approche par classification de scores [171] cherche à prendre une décision à partir du vecteur de scores de dimension N . Le but des méthodes de fusion basées sur des classificateurs est de chercher une séparatrice entre les 2 classes, Client et Imposteur, dans un espace des scores à N dimensions. La séparatrice peut être plus ou moins complexe selon le classificateur utilisé. Toutes les méthodes classiques de reconnaissance de formes et d'apprentissage automatique peuvent être utilisées pour la fusion des scores qui est un problème de classification commun dans le domaine des systèmes multi-algorithmes, dont l'objectif est d'utiliser un classificateur pour combiner plusieurs systèmes sources (ici ce sont des systèmes biométriques que l'on veut fusionner) afin de consolider la décision de classification. Un grand nombre de méthodes ont été utilisées, celles utilisées dans notre travail sont :

- Les machines à vecteurs de support (SVM).

3.3.5 Description de la reconnaissance d'expressions faciales [131]

La phase qui suit la détection du visage est la reconnaissance des expressions faciales qui fait également l'objet de plusieurs travaux dans le domaine de la vision par ordinateur [72]. La reconnaissance automatique des expressions faciales se compose de deux étapes qui sont la caractérisation et la comparaison ou classification. La première étape extrait un jeu de primitives pertinentes et discriminantes, autrement dit, c'est la création d'un ensemble de caractéristiques qui reflètent convenablement les informations décrivant l'expression manifestée par le visage. Au cours de la deuxième étape, à savoir la comparaison ou la classification, il y a recours aux techniques d'évaluation de relation pour faire soit une mesure de similarité (ressemblance), soit une mesure de distance (divergence) telles que les distances euclidiennes, Mahalanobis, ...

Comme tout système de biométrie automatique, un système de reconnaissance automatique des expressions faciales fonctionne en deux modes [172] : mode de vérification dit 1 contre 1 où il compare le vecteur requête de caractéristiques d'une expression prédite avec son modèle présent

dans la base. La sortie de ce mode est « vraie » en cas de ressemblance ou « faux » en cas d'échec. Le deuxième genre de fonctionnement est un mode d'identification dit 1 contre N où le système réalise une comparaison entre les caractéristiques d'une expression en cours de test avec N vecteurs (N égale à sept) stockés dans la base. Le résultat dans ce cas est l'identification de l'expression ou un message indiquant l'échec de reconnaissance.

3.3.5.1 *Techniques existantes* [31]

Selon Albert Mehrabian [32], l'expression faciale d'un interlocuteur compte pour 55 % de l'effet d'un message oral et l'intonation de la voix contribue à 38 % alors que la partie verbale induit seulement 7 % d'effet. De ce constat se sont inspirés plusieurs travaux qui se sont orientés vers l'étude des expressions faciales notamment la reconnaissance et la reconstruction.

La littérature présente un nombre important de techniques qui traitent la reconnaissance des expressions faciales, toutefois nous nous intéressons seulement à celles qui sont toujours présentes dans les travaux de recherches. Avant d'aller plus loin, il est essentiel de noter que plusieurs techniques traitent à la fois la détection de visage et ses composants et la reconnaissance de visage et d'expressions faciales.

➤ *La reconnaissance par visages propres :*

C'est la première méthode utilisée dans la biométrie automatique (connue sous le nom « Eigenfaces » en anglais). Elle est connue également sous le nom d'Analyse en Composantes Principale (ACP). Huang et Huang [173] firent le constat que l'étude de 44 unités d'actions (AUs) n'est pas facile et ils proposèrent de les remplacer par 10 paramètres d'actions (APs). Ils enlevèrent les 10 APs en se basant sur la différence entre un visage neutre et un autre expressif et appliquèrent un ACP sur ces derniers pour réduire la dimension à deux et simplifier le processus de reconnaissance.

Les auteurs utilisèrent un classificateur de distance minimale pour le traitement des six expressions universelles en calculant le profil de corrélations des APs. Le taux le plus élevé des corrélations fixe la classification finale de l'expression examinée. Le système élaboré arrive à reconnaître les expressions étudiées et présente un taux de reconnaissance égal à 84.5 % [130].

➤ *La reconnaissance par « Fiserfaces » :*

En 1997, Belhumeur et al. [45] ont proposé une alternative à l'approche « eigenfaces » sous le nom « Fiserfaces ». Cette technique est connue également sous le nom d'Analyse Discriminante Linéaire (ADL). La méthode ADL tient compte, en général, du problème de séparation des classes liées à l'ACP.

Les travaux [174] et [175] de Lyons et al. ont présenté un système pour classifier des images du visage en catégories selon trois critères, à savoir sexe, race et expression faciale. Nous nous limitons à la partie qui traite les expressions faciales, le système proposé combine les avantages des graphes étiquetés par des réponses de filtres de Gabor à ceux des algorithmes ACP et ADL. Ils ont utilisé six classificateurs binaires pour classifier les sept expressions faciales. L'apprentissage de chaque classificateur a été fait indépendamment des autres qui sont par la suite combinés pour produire un reconnaiseur. Ce dernier utilise une distance normalisée par rapport au centre de groupe comme un

facteur de décision pour déterminer la classe finale d'une image d'entrée et affecte l'expression neutre dans le cas d'échec de la classification. Le taux de reconnaissance du système atteint 92 % lors de tests avec 193 images de neuf femmes japonaises (sujets familiers) et décroît jusqu'à 75 % lors de test avec des sujets non familiers [130 ,174].

➤ *La reconnaissance par SVM :*

La technique des machines à vecteurs de support, connue sous le nom de SVM, est une approche discriminante d'apprentissage statistique supervisé. Elle a un avantage majeur, à savoir, la classification linéaire des problèmes qui ne sont pas linéairement séparables.

Valstar et al. ont proposé dans [176] une approche de fusion entre différents niveaux d'abstraction de types de caractéristiques faciales, de schémas de classification et diverses règles de fusion. Ils implémentèrent trois stratégies de fusion et utilisèrent un classificateur multi-classe dit GentleSVM qui se base sur un noyau sigmoïde et ils traitèrent le problème de reconnaissance du sourire spontané par opposition du sourire sournois. Ils testèrent leurs trois approches sur 102 vidéos de sourires spontanés et 100 vidéos avec sourires non spontanés et montrèrent que le meilleur taux de reconnaissance est égal à 94 %, qui est atteint par la troisième modalité de fusion.

➤ *La reconnaissance par réseaux de neurones :*

D'après Haykin [177], un réseau de neurones ressemble au cerveau en deux points : l'acquisition de la connaissance est le résultat d'un processus d'apprentissage et de mémorisation d'une connaissance prenant fait des poids des connections entre les neurones. L'avantage de la technique des réseaux de neurones réside dans sa capacité d'apprendre, de mémoriser et de généraliser pour mieux classifier de nouvelles données.

Zhang et al. [178] utilisent les réseaux de neurones pour le problème de reconnaissance des expressions faciales. La base d'images utilisée pour les expériences contient 213 images féminines avec expressions faciales dont le nombre d'image par expression est en grosso modo le même. La performance de reconnaissance des sept expressions par le système atteint 90.1% d'après [130].

➤ *La reconnaissance par Modèles de Markov cachés :*

Les modèles de Markov cachés connus sous l'appellation anglaise (HMM : Hidden Markov Models) ont été bien utilisés dans les applications de reconnaissances de paroles et de caractères depuis de nombreuses années, ils ont été ensuite appliqués à la reconnaissance de visages et d'expressions faciales. Cohen et al. [179] ont proposé une méthode multi-niveaux basée sur les HMMS pour la segmentation et la reconnaissance des expressions faciales à partir d'une séquence vidéo. Le premier niveau de l'architecture proposée est composé de six HMMs indépendants, un HMM pour chacune des expressions universelles. Les caractéristiques extraites du visage correspondent aux entrées des HMMs de bas niveau et la séquence des sorties de ces derniers est codée puis utilisée comme vecteur d'observation du modèle de haut niveau. Cohen et al. ont montré que la performance de leur approche atteint un taux de reconnaissance environ 82.46 % pour une personne familière et un taux d'environ 58.63 % pour une personne non familière.

➤ *La reconnaissance par Modèle Actif d'Apparence (MAA) :*

La technique MAA a été proposée en 2001 par Cootes et al. [180]. La méthode permet la détermination conjointe des variations de forme et de texture dans un ensemble d'apprentissage.

Parmi les travaux qui se sont intéressés à la reconnaissance des expressions par les modèles actifs d'apparences, nous citons le travail de Franck Davoine et al. [181] qui propose une solution qui reconnaît l'expression d'un visage par discrimination linéaire. Nous nous sommes intéressés en particulier aux travaux traitant la reconnaissance d'expressions faciales et utilisant une technique parmi les plus présentes dans la littérature. Néanmoins, il existe d'autres techniques telles que celles décrites dans [182] et [183] basées sur le calcul du flux optique représentant la direction et l'importance d'un mouvement. Ces deux derniers travaux présentent un taux de reconnaissance des six expressions universelles égal à 88% d'après [130].

Nous trouvons également des techniques qui utilisent la théorie de l'évidence basée, soit sur les traits permanents [184], soit sur les traits transitoires [130] du visage et qui atteignent les 64.5% en moyenne pour la reconnaissance de la joie, les expressions considérées étant « joie ou dégoût » et « surpris ou peur ».

De même, il existe des techniques 3D, à titre d'exemple [185] présente une approche basée sur des « patches » extraits de visages 3D et donnant un taux de reconnaissance des six expressions égales à 97.75 % et ce, en utilisant des techniques de multiboosting tels que MultiBoost avec un noyau LDA ou noyau des plus proches voisins (KNN : K-Nearest Neighbor) ou noyau naïf de Bayes (NB). La même approche basée sur la méthode SVM avec un noyau linéaire atteint une performance égale à 98.81% pour la reconnaissance des mêmes expressions.

3.4 Conclusion

Dans ce chapitre, nous avons étudié et décrit la reconnaissance automatique de visages en 2D, et plusieurs techniques ont été proposées ainsi en décrivant leur principe de fonctionnement pour chaque procédé.

Nous avons également énuméré les limitations des systèmes biométriques lorsqu'ils n'utilisent qu'une seule modalité biométrique et les différentes multimodalités possibles. Nous avons achevé ce chapitre par une description de la reconnaissance d'expressions faciales ainsi que les travaux relatifs, étape essentielle pour la définition et l'élaboration d'un système de reconnaissance tridimensionnelle du visage à présenter dans le chapitre 4.

Les résultats expérimentaux liés à notre système de reconnaissance bidimensionnelle automatique du visage seront présentés à l'occasion du chapitre 5.

Dans le chapitre suivant, nous allons détailler principalement les techniques à utiliser dans chacune des phases de notre proposition pour la reconnaissance tridimensionnelle du visage.

Notre méthode pour la vérification des visages 3D

4.1 Introduction

La reconnaissance automatique des individus en fonction de leurs visages a de nombreuses applications utiles, en particulier dans des domaines tels que l'interaction homme-machine (robotique intelligente dans des environnements sociaux) et la sécurité (surveillance et contrôle d'accès). Les nombreux avantages possibles poussent les chercheurs à améliorer la fiabilité des systèmes d'authentification basés sur l'empreinte digitale du visage. On peut citer l'acceptation de cette technologie par le public, l'absence de contact direct ainsi que le faible coût de mise en œuvre. Malheureusement, la présence d'une expression faciale complique la tâche et trouble l'adoption industrielle de cette approche. Ce chapitre présente donc la reconnaissance 3D de visage ainsi la reconnaissance multimodale de visage par la fusion multi-algorithmes, nous allons présenter deux contributions distinctes.

Notre première contribution proposée permet, entre autres, d'augmenter la performance de la biométrie faciale en offrant un système de reconnaissance d'expressions faciales. Dans la première section, on aura la chaîne d'exécution de notre approche tout en décrivant ses différentes étapes, nous expliquerons le fondement théorique de notre technique de reconnaissance basée sur MB_LBP. Dans la deuxième section, nous exposerons le capteur Kinect en citant ses caractéristiques et son mode de fonctionnement. Au cours de la dernière section, nous décrirons en détail les méthodes de reconstruction et d'animation 3D du visage en exploitant le capteur Kinect.

La deuxième contribution basée sur la reconnaissance de visage 3D en utilisant multi-algorithmes de visages qui est testée sur une base de données contrôlée FRGC v1.0. Puisque les taux d'erreurs associés à des systèmes biométriques restent relativement élevés lorsque les changements d'éclairage, d'expressions et de rotation de la tête sont pris en compte. Pour pallier ces inconvénients, une solution est l'utilisation de l'image 3D en utilisant cette base de données.

4.2 Première contribution : *La reconnaissance de visage en 3D liées aux expressions faciales basée sur la méthode MB_LBP.*

Nombreuses sont les extensions effectuées dans les systèmes et les techniques d'acquisition d'images 2D et vidéo pour perfectionner les approches d'analyse de visage telles que la détection, le suivi, la reconnaissance et la synthèse. Ces développements améliorent de manière typique l'identification d'une expression faciale. Cependant, le succès de traitement des expressions faciales

Chapitre 4: Notre méthode pour la vérification des visages 3D

est encore pénible à cause de certains problèmes tels que les variations des conditions d'éclairage, les changements de pose de la tête et les occlusions partielles.

A l'avenir des caméras 3D, une nouvelle modalité faciale a surgi dans le domaine de vision par ordinateur. C'est la géométrie 3D du visage. Grâce à ceci, l'acquisition de données est plus complète ainsi l'analyse de visage a pu faire face aux modifications de pose et d'illumination. Néanmoins, le prix de ces scanners est très cher et leur configuration est délicate.

D'un autre côté, nous partageons avec Weise et al [186] que la création des créateurs numériques exige une fusion d'un scan 3D, d'une acquisition d'apparence et d'une capture de mouvement. Cette fusion est complexe ainsi, elle nécessite un pré et post traitement manuel substantiel. En conséquence, la création d'un système d'animation d'un modèle virtuel de haute qualité nécessite de longs délais et des coûts de production importants.

Pour produire un système capable d'identifier une expression faciale, nous proposons de tirer profit de la performance atteinte par les techniques 2D et de corriger les lacunes par l'exploitation de la géométrie 3D du visage grâce à la carte de profondeur offerte par le capteur Kinect.

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

L'architecture de notre méthode est composée de sept étapes : l'acquisition de données suivie d'un prétraitement puis la détection du visage. Ensuite, nous avons l'étape quatre de caractérisation suivie de l'étape cinq ou six. L'étape cinq sert à apprendre les caractéristiques des expressions sa étudiées par contre, l'étape six permet la reconnaissance d'une expression de test.

4.2.1 Description des étapes de l'approche proposée

Le problème à résoudre avec notre approche se résume à l'implémentation d'un système capable de détecter le visage d'une personne avec une vue frontale puis reconnaître son expression faciale.

Pour atteindre notre objectif, il est indispensable d'avoir une base d'images de toutes les expressions reconnues par le système et un modèle générique qui peut présenter toutes ces expressions.

La figure 4.1 ci-dessous présente la chaîne d'exécution de notre système **AREFK** (Automatic Recognition Expression of Face with Kinect) développé.

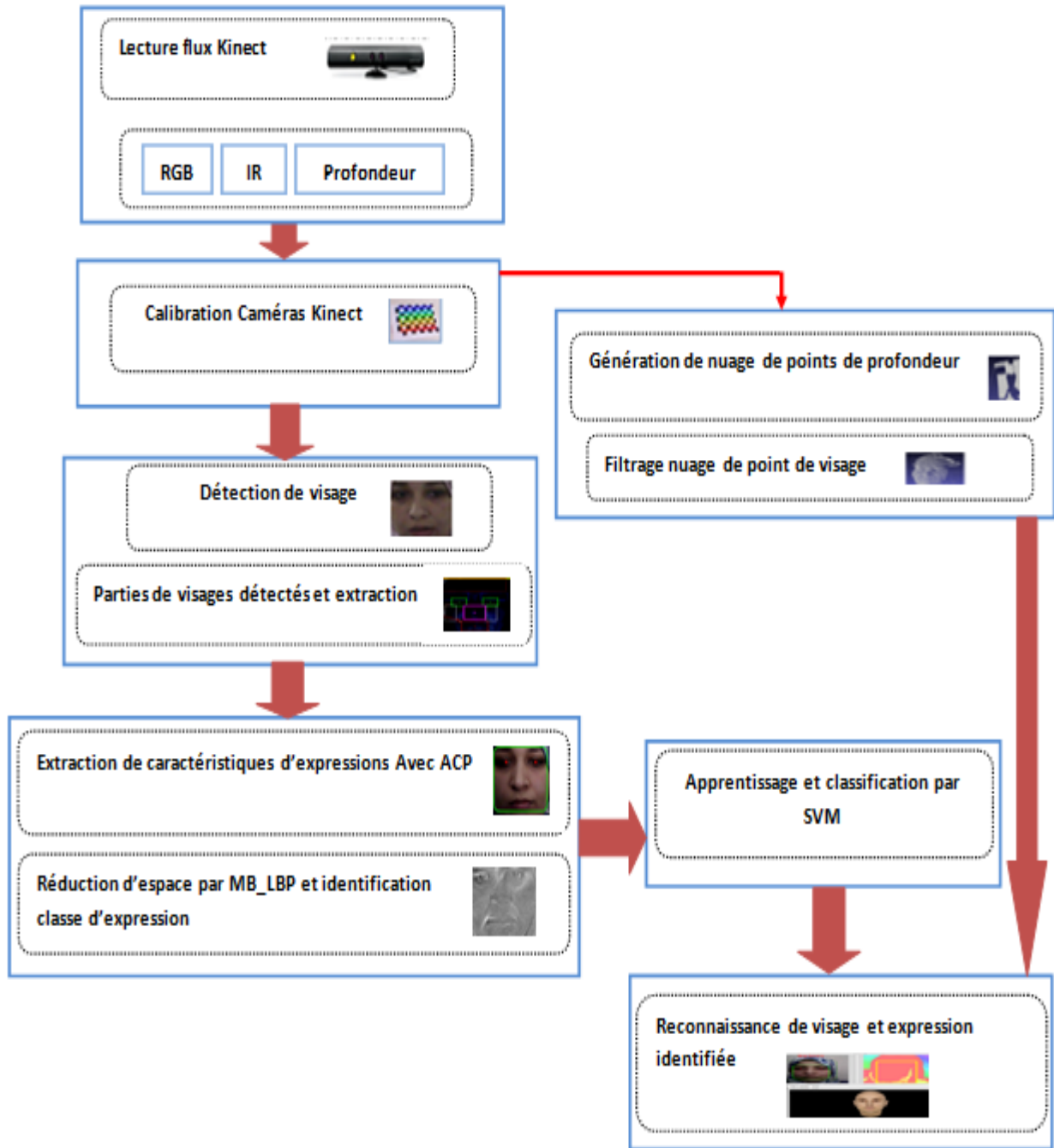


Figure 4.1. Système automatique de reconnaissance des expressions de visages en 3D en utilisant la caméra Kinect (AREFK).

Dans ce qui suit, nous allons détailler chacune de ces étapes.

4.2.1.1 Acquisition de données

L'acquisition des données est la première étape de notre système permet d'acquérir des données à partir du monde réel. Autrement dit, cette étape sert à transformer une observation d'une personne (ou objet) physique effectuée par un capteur en un flux utilisable dans l'étape suivante.

Dans notre cas, toutes les données d'entrées sont acquises via le système Kinect (figure 4.2). Ce dernier tolère des captures simultanées d'une image couleur 2D et d'une carte de profondeur 3D.

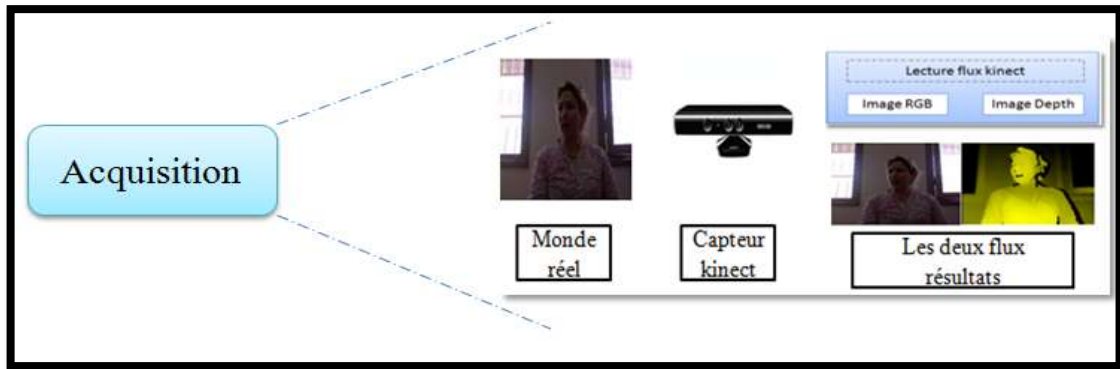


Figure 4.2. Détails de l'étape « Acquisition ».

4.2.1.2 Prétraitement

L'étape de prétraitement s'avère nécessaire pour l'amélioration de la qualité des données acquises. Elle se décompose dans notre système en deux tâches (figure 4.3) : le calibrage des deux caméras du Kinect et le filtrage de flux de profondeur pour représenter seulement les informations utiles, c'est-à-dire garder seulement les données représentant le visage.

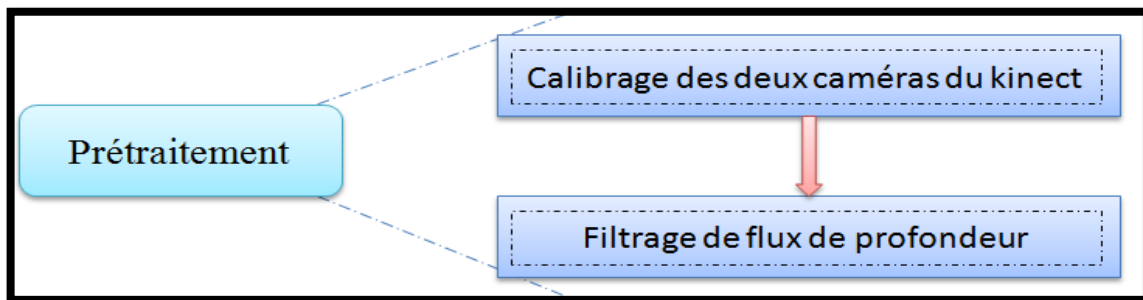


Figure 4.3. Détails de l'étape « Prétraitement ».

Après une normalisation de l'image et une mise à l'échelle vers une taille fixe (ex. : $m*n$), l'image du visage peut être considérée comme un vecteur dans un espace multi-dimensionnel R^{mn} . Ce vecteur est obtenu en mettant, simplement, dans un ordre fixe les éléments de la matrice des pixels. Afin de définir une image (ou un point) dans l'espace des images, nous devons spécifier une valeur pour chaque pixel de cette image. Le nombre de points constituant cet espace devient rapidement très grand, même pour les images de petite dimension. Cette dimensionnalité pose un certain nombre de problèmes pour les algorithmes de reconnaissance, qui se basent sur cette représentation de l'image, à savoir : [142]

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

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

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

4.2.1.3 Détection

L'étape de détection du visage est une étape primordiale permettant de chercher la présence d'un visage en vue frontale dans la scène capturée. Une fois détecté, on doit récupérer sa localisation et suivre son mouvement.

Étant donné que nous traitons des séquences vidéo, la phase de détection implique implicitement la phase de suivi du visage dans la scène, puisque nous traitons la séquence vidéo image par image. Cette étape se décompose de trois tâches à savoir, la détection du visage à l'aide de la méthode « Viola&Jones », l'extraction des traits faciaux (sourcils, yeux, nez, bouche, ...) et le suivi des déplacements du visage dans la scène (Figure 4.4).

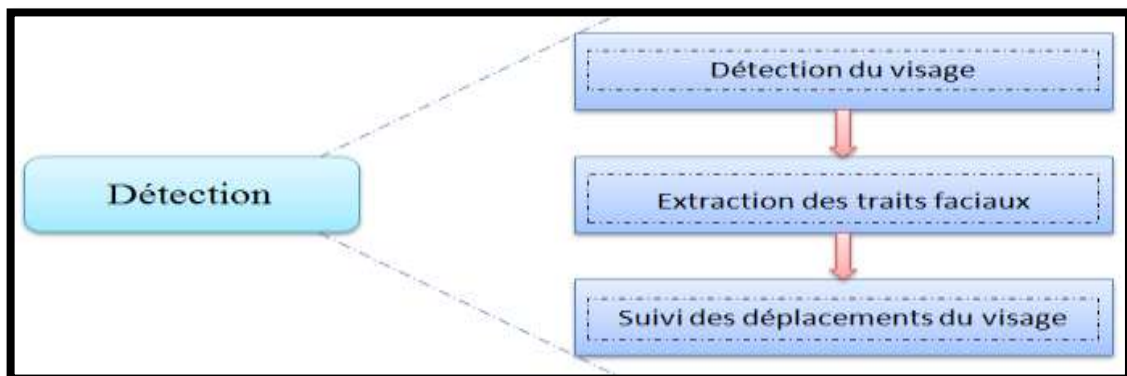


Figure 4.4. Détails de l'étape « Détection ».

➤ Méthode de Viola et Jones

La méthode « Viola&Jones » a été proposée au départ pour la détection de visages dans une image numérique ou séquence vidéo puis utilisée pour détecter d'autres objets comme les voitures... La bibliothèque OpenCV présente une implémentation de cette méthode sous le nom « détecteur en cascades de Haar ». Le point fort de cette méthode est la rapidité de détection ce qui la rend capable de s'exécuter en temps réel et de répondre aux exigences du traitement vidéo. Toutefois, elle présente quelques limites telles que la difficulté de détection simultanée de plusieurs vues de même objet, la durée nécessaire à la phase d'apprentissage des cascades est relativement assez grande et le nombre d'échantillons d'apprentissage est important.

Pour évaluer la performance de notre système de détection, nous utilisons les termes suivants :

- Vrai positif : le système détecte la zone du visage dans l'image et il retourne ses coordonnées (position, hauteur et largeur) comme résultat.
- Faux positif : le système détecte une zone de l'image ne contenant pas de visage et il retourne ses coordonnées.
- Vrai négatif : le système a bien détecté que l'image ne présente pas un visage.
- Faux négatif : l'image présente un visage par contre le système n'arrive pas à le détecter.

4.2.1.4. Caractérisation

L'étape de caractérisation consiste à dégager les caractéristiques pertinentes de l'expression faciale étudiée et d'éliminer les informations redondantes.

Puisque nous faisons usage de la méthode « eigenface », la première tâche de cette étape consiste à générer un ensemble de visages propres dont le nombre est égal au nombre d'images de la base. La deuxième tâche est le maintien des X premiers visages propres générés qui disposent de l'information caractéristique de visages. La dernière tâche est le calcul de la matrice des poids dans l'espace engendré par les visages propres retenus. (Figure 4.5)

Pour la réduction d'espace engendré des images d'apprentissage, nous faisons l'usage de la méthode « MB_LBP ».

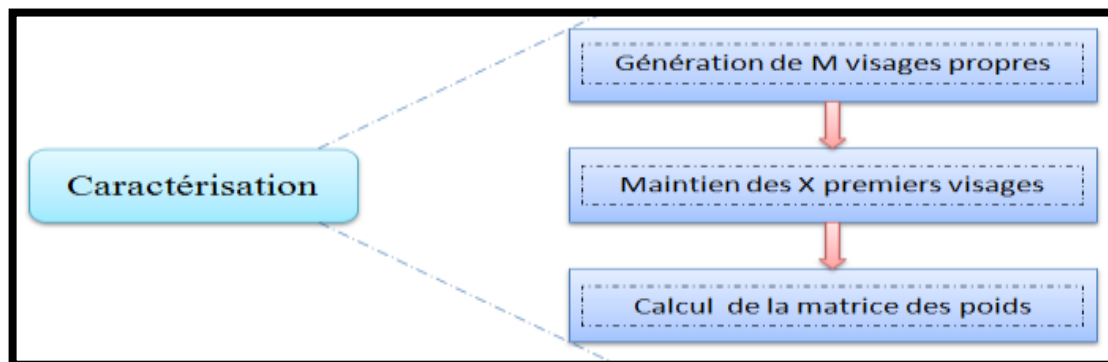


Figure 4.5. Détails de l'étape « Caractérisation ».

➤ La réduction d'espace

A. Multi Block LBP, MB-LBP (MB_LBP)

Afin de saisir non seulement les microstructures, mais aussi les macrostructures, Li et al. [148], [148] ont proposé une LBP multi-bloc (multiblock LBP, MB-LBP), qui, au lieu de comparer les pixels, compare les intensités moyennes des sous-régions voisines. La LBP originale peut être considérée comme un cas particulier de la MB-LBP. La figure 4.6 montre un exemple de MB-LBP, où chaque sous-région est composée de six pixels.

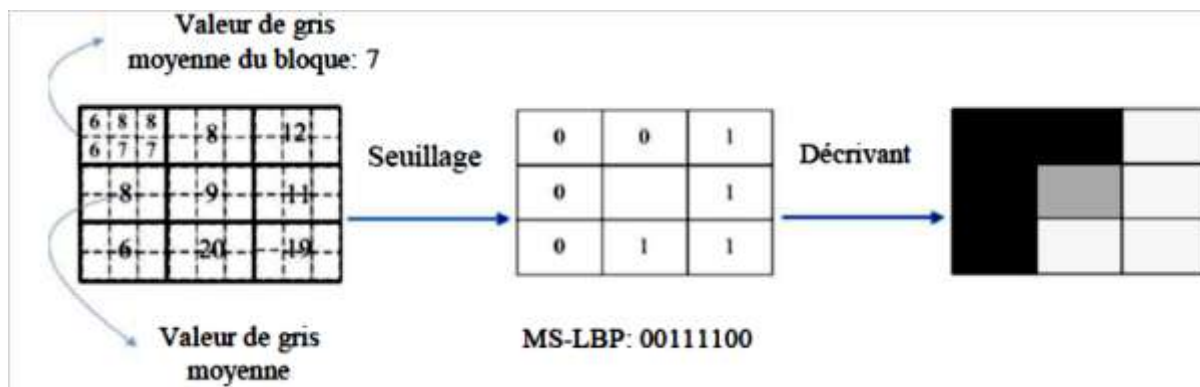


Figure 4.6. Exemple de la méthode MB-LBP.



Figure 4.7. Exemple de la méthode MB_LBP pour un visage.

L'adoption de cette approche (MB_LBP) nous semble un bon compromis entre performance et complexité algorithmique. L'objectif de cette méthode de réduction d'espace est de trouver une base réduite de projections qui optimise un certain critère et l'espace original est ensuite projeté sur cette base de dimension inférieure. De ce fait, on réduit considérablement la complexité des calculs tout en conservant l'information discriminante entre les images. La classification est habituellement effectuée selon une mesure de distance simple dans l'espace multidimensionnel final en utilisant le SVM.

4.2.1.5. Apprentissage

Notre système exige une phase d'apprentissage, il y aura une collecte des informations sur les expressions à identifier et leur sauvegarde sous forme de descripteurs faciles à utiliser.

Dans notre cas, un descripteur est un vecteur de poids qui décrit les poids de chaque image dans l'espace de visages propres. Nous avons retenu seulement X visages propres parmi ceux générés donc nous aurons X vecteurs de poids qui seront sauvegardés sous forme d'une matrice (Figure 4.8) dite matrice des poids en utilisant la méthode SVM.



Figure 4.8. Détail de l'étape « Apprentissage ».

➤ Pourquoi choisir les SVM? [187]

Plusieurs facteurs sont pris en considération lors du choix d'une méthode pour concevoir un système de reconnaissance de visage, mais les plus importants d'entre eux restent :

Chapitre 4: Notre méthode pour la vérification des visages 3D

- Le coût élevé et la difficulté trouvés lors de la conception d'un système de reconnaissance de visages : Il s'agit là de la difficulté rencontrée lors de l'implémentation des algorithmes de reconnaissance ainsi que le coût élevé des bases de données et des informations préalables.
- Les performances des méthodes de reconnaissance : Les performances d'un système de reconnaissance de visages sont calculées selon deux facteurs importants : la précision de la reconnaissance ainsi que le temps nécessaire pour le mettre au point.

Comme nous l'avons dit précédemment, la méthode SVM est une méthode qui fait partie de la gamme des méthodes de l'apprentissage statistique. Le système de reconnaissance est conçu avec des images de formation et aucune connaissance préalable n'est requise sur l'image.

Ces méthodes ont la particularité d'être plus simples à mettre en œuvre que les méthodes basées sur la géométrie. L'algorithme de reconnaissance est plus simple et il y a beaucoup moins de lignes de codes que dans un algorithme basé sur d'autres approches. Les images de formation nécessaires pour l'implémentation du système de reconnaissance ne posent pas de problèmes puisque beaucoup de bases de données sont fournies gratuitement dans le domaine de la recherche.

L'inconvénient des méthodes basées sur l'image c'est qu'elles sont moins performantes que les méthodes géométriques, cela est dû surtout au nombre d'images de formation qui est relativement petit par rapport au nombre d'images total de visage et non-visage qui existe.

L'utilisation d'un très grand nombre d'images de formation va entraîner un système de détection très lent.

La méthode SVM agit directement sur la précision en minimisant l'erreur de généralisation. De plus, le système ne sera représenté que par un petit nombre de données qui représentent la frontière entre la classe de visage recherché et les autres classes (le concept de Vapnik). Qui dit moins de données dira moins de temps lors de la décision. Les deux derniers points nous permettent de penser que le système de reconnaissance basé sur SVM est plus performant que les systèmes implémentés avec les autres méthodes faisant partie de l'approche image.

L'un des autres points forts de la méthode SVM c'est qu'on peut facilement l'implémenter avec une méthode d'extraction de paramètres telle que la ACP. Cela va nous permettre de réduire la dimension de l'image et de gagner du temps lors de la reconnaissance. De plus, la décorrélation des données fait que la séparation des deux classes sera plus robuste. On déduit de tout cela que du point de vue théorique, un système de reconnaissance basé sur la méthode SVM semble être facile à concevoir, et n'est pas gourmande en matière de temps et précision de la reconnaissance.

Dans cette partie, nous allons faire l'étude théorique de la méthode SVM en détail. Nous allons voir comment arriver à formuler mathématiquement le problème de l'hyperplan optimal dans les deux cas linéairement séparables et non linéairement séparables.

Les SVM constituent une classe d'algorithmes basée sur le principe de minimisation du « risque structurel » décrit par la théorie de l'apprentissage statistique de Vapnik et Chervonenkis qui utilise la séparation linéaire. Cela consiste à séparer par hyperplan des individus représentés dans un espace de dimension égale au nombre de caractéristiques, les individus étant alors séparés en deux

classes. Cela est possible quand les données à classer sont linéairement séparables. Dans le cas contraire, les données seront projetées sur un espace de plus grande dimension afin qu'elles deviennent linéairement séparables.

4.2.1.6. Reconnaissance

Au cours de la phase de reconnaissance, il y aura une comparaison entre les caractéristiques acquises de l'expression requête avec celles issues de l'apprentissage. C'est une comparaison du vecteur de poids test par rapport aux informations sauvegardées dans la matrice des poids. L'objectif de cette étape est la détermination qu'une expression existe bien dans notre référence. Pour le cas de notre approche (figure 4.9), nous avons recourt à la distance euclidienne comme étant comparateur entre le vecteur requête et la référence. Avoir une distance minimale entre le vecteur test et une parmi les colonnes de la matrice de poids indique que l'image correspondante à ce dernier vecteur est la plus proche de l'image test et son identité est celle de l'expression de test.

Le principe de calcul de la distance euclidienne est présenté dans le chapitre 3, dans l'étape de reconnaissance.

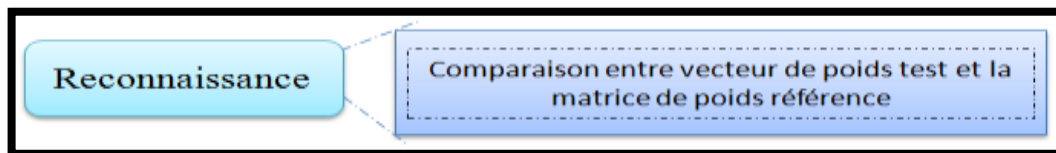


Figure 4.9. Détail de l'étape « Reconnaissance ».

4.2.2. Présentation du Kinect [131]

4.2.2.1 Caractéristiques du Kinect

Le capteur Kinect a été lancé en novembre 2010, était destiné pour être un dispositif d'entrée de capture de mouvement pour le XBOX 360. C'est un parmi les outils d'interaction naturelle permettant à l'utilisateur de contrôler des jeux vidéo sans manette, mais par l'intermédiaire des gestes (à titre d'exemple lever la main pour changer la couleur d'un bouton dans une interface graphique et/ou pour quitter l'application) et des commandes parlées (exemple dire « exit » pour arrêter votre application). Ce capteur se constitue de cinq composants clés qui sont une série de microphones, une caméra RGB, une caméra de profondeur 3D composée d'une source de lumière IR et un capteur IR, un moteur d'inclinaison et un accéléromètre à trois axes (figure 4.10).

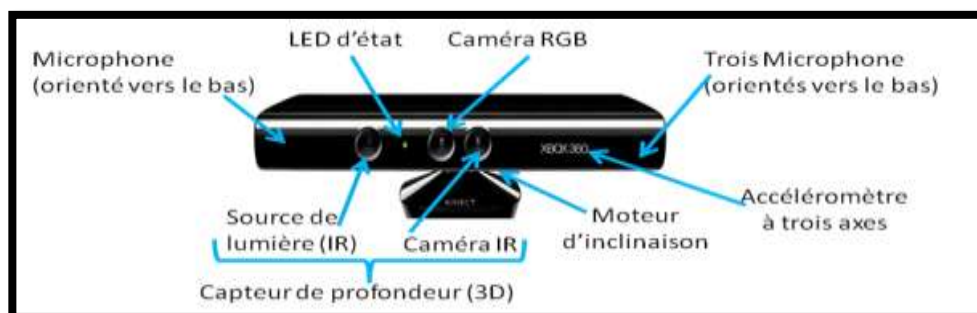


Figure 4.10. Le capteur Kinect de Microsoft.

4.2.2.2. *Fonctionnement*

Le Kinect permet de capturer et visualiser simultanément une image couleur RGB2D et une carte de profondeur 3D. Le capteur de profondeur du Kinect est composé d'une source laser de lumière infrarouge qui projette un pattern de points et d'une caméra IR qui récupère ce dernier. La distorsion entre le pattern émis et celui reçu est utilisée pour reconstruire la valeur de chaque rayon réfléchi. Le capteur de profondeur convertit les intensités en distance et génère une grille de 640 x 480 en une résolution VGA contenant des valeurs sur 11 bits (ce qui permet de fournir 2.048 niveaux de sensibilités) et ceci à une fréquence moyenne de 30 frames par seconde. La résolution de la dimension de profondeur (selon l'axe z) est en centimètres par contre la résolution spatiale (selon les axes x et y) est en millimètres.

L'image couleur (ou image RGB) est fournie dans la même résolution et fréquence que le flux de profondeur. Toutefois, les deux flux ne se correspondent pas naturellement à cause des différents paramètres intrinsèques et extrinsèques des deux caméras du Kinect (caméra RGB et IR). Ceci rend la phase de calibrage une étape primordiale afin d'aligner les deux signaux.

4.2.2.3. *Calibrage des caméras*

L'objectif du calibrage d'une caméra (respectivement des caméras) est la détermination de ses (respectivement leurs) paramètres intrinsèques et extrinsèques. Les paramètres intrinsèques sont les distances focales horizontales et verticales, l'angle entre les lignes et les colonnes de l'image (souvent proche de $\frac{\pi}{2}$) et les coordonnées de l'intersection de l'axe optique avec le plan image. Les paramètres extrinsèques représentent les positions et orientations relatives à l'objet du monde réel et à la caméra. En cas de deux caméras ou plus, ces paramètres incluent en surplus la position et l'orientation de chaque caméra par rapport à l'autre [189, 188].

La phase de calibrage se compose de deux sous-étapes, à savoir la détermination de la fonction de projection de chaque caméra et l'extraction des paramètres. La fonction de projection P est la fonction qui associe à chaque point 3D un point 2D dans l'image et elle est formée dans le plan rétinien de la caméra. Après avoir calculé P, les paramètres intrinsèques et extrinsèques peuvent être extraits facilement [190].

Afin d'obtenir la corrélation désirée par le calibrage, on commence par réaliser une prise d'un nombre d'images relatives à un objet dont sa structure est parfaitement connue. Cet objet présente la mire de calibrage ; à titre d'exemple, le damier dans le cas d'OpenCV (figure 4.11).

Les positions des coins d'intersection entre les lignes et les colonnes sont connues avec précision dans le repère attaché à la mire. Les positions de ces coins dans l'image sont extraites par un traitement d'image. Une fois ces dernières sont déterminées, il ne reste que l'établissement de lien entre le coin 3D du monde réel avec sa projection dans l'image. Comme résultat on obtient la matrice de projection P [190].

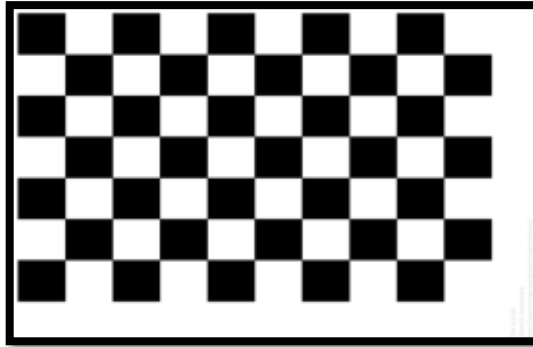


Figure 4.11. Mire de calibration d'OpenCV.

En cas de vision stéréoscopique, une fois qu'on a les caractéristiques de chaque caméra ainsi que sa position par rapport à l'autre, reste à réaliser la rectification des deux images par leurs projections sur un plan parallèle à la droite ajustant les centres des caméras. Cette rectification permet de réduire énormément le temps de calcul lors de la corrélation. La corrélation étant l'étape de mise en correspondance consistant à calculer la carte de disparité mesurant la distance entre deux points dans les deux images.

Il est essentiel de noter que ce processus est semi-automatique dans la plupart des cas. Néanmoins, la bibliothèque OpenNI propose une fonction qui permet d'aligner le flux RGB et le flux de profondeur du Kinect.

4.3 Reconstruction et animation 3D des expressions faciales [131]

Dans la suite nous citons les techniques de la reconstruction et animation 3D des expressions faciales qui nous intéressent dans nos futurs travaux liés à notre domaine de recherche ;

4.3.1. Description de la reconstruction 3D

La reconstruction 3D consiste à retrouver la géométrie de la scène observée par le système de vision dans ses dimensions réelles. Cette reconstruction se compose de trois sous étapes : calcul du nuage de points à 3D, interpolation de ces points par un algorithme de maillages parmi ceux disponibles pour récupérer une forme présentant une continuité et la dernière sous étape est la texturation de la forme obtenue.

Plusieurs méthodes de reconstruction 3D sont présentes dans l'état de l'art, reconstruction par mono-caméra (géométrie à partir d'obscurité (Shape from shading), géométrie à partir de centre et géométrie par stéréo-photométrie), reconstruction par vision active (lumière structurée et laser) et reconstruction par multi-caméra et vision passive (stéréoscopique, sculpture spatiale (space carving), cônes généralisés et silhouette) [191]. Nous nous intéressons uniquement au troisième type des méthodes vu que nous avons utilisé le capteur Kinect de Microsoft composé de deux caméras vidéo et profondeur.

4.3.2. Vision stéréoscopique

Chez l'être humain, la vision permet de distinguer trois sortes d'éléments qui sont les formes, les reliefs et les couleurs. Nous nous intéressons au cas particulier de la vision des reliefs et des distances connues également par le nom de vision stéréoscopique, cette vision est possible grâce à

Chapitre 4: Notre méthode pour la vérification des visages 3D

la capacité de cerveau d'intégrer et d'analyser les deux images légèrement différentes (deux points de vue : droite et gauche) fournies par chaque œil. Le résultat fourni par ce type de vision est la synthèse de la perception spatiale [192].

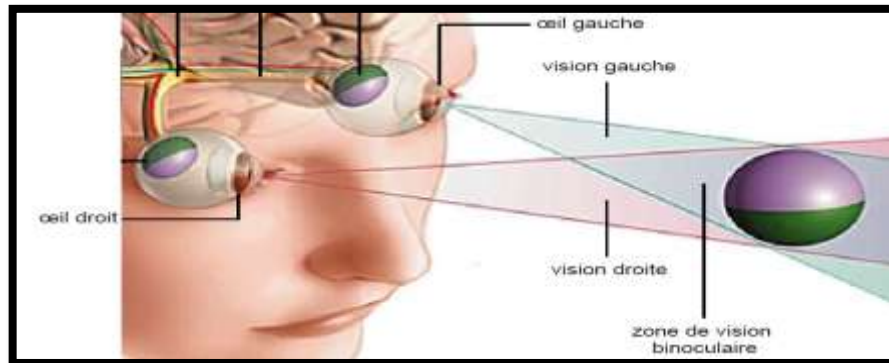


Figure 4.12. Vision humaine stéréoscopique d'après [192].

En conséquence, la stéréovision par ordinateur est l'imitation de la conception de relief naturellement présente chez l'homme afin d'ajouter une troisième dimension à la vision classique, cette dimension est la profondeur.

Le but principal de la stéréoscopie est la reconstruction d'une scène 3D à partir de deux images du même monde réel prises de points de vue différents en même instant. Néanmoins, la difficulté à atteindre cet objectif réside à trouver les points de correspondance entre les images [188]. Autrement dit, déterminer les points qui présentent la projection du même point matériel dans les images acquises. Ce besoin rend le calibrage des caméras une phase fondamentale pour la reconstruction 3D, ainsi que la rectification d'une image par rapport à l'autre et la mise en correspondance entre les pixels des deux images.

4.3.3. Description de l'animation 3D

Le terme animation relatif au domaine traitement d'images, désigne un procédé permettant d'obtenir des images animées. Ce dernier consiste à donner l'illusion du mouvement à l'aide d'une suite d'images et à synchroniser inter-flux et intra-flux (Hachette, 1997). Dans le contexte d'une scène 3D, nous gardons la même définition tout en ajoutant le passage d'une forme initiale vers une autre finale dans une durée de temps déterminée.

Nous nous intéressons en particulier, à l'animation 3D de visage qui permet de changer l'expression faciale présentée par un modèle virtuel d'une classe vers une autre. Parmi les méthodes d'animations très présentes dans l'état de l'art nous signalons l'animation d'images clés ou Key-frame (connue également sous le nom d'interpolation géométrique) et l'animation par paramétrisation [186].

4.3.4. Techniques existantes

La reconstruction des expressions faciales est étudiée dans le domaine 2D par Abboud et al. [193]. Ces derniers sont implémentés deux approches pour la synthèse des expressions faciales en utilisant le modèle actif d'apparences : une première modélisation basée directement sur le modèle proposé par [180] et une deuxième basée sur un modèle évolutif. Dans les deux implémentations, ils

Chapitre 4: Notre méthode pour la vérification des visages 3D

annulent en premier lieu l'expression d'un visage quelconque, puis ils calculent une expression artificielle sur ce même visage à partir d'une seule image.

La reconstruction d'un modèle de visage 3D est examinée par Kemelmacher-Shlizerman et al [194]. Ces auteurs ont proposé une méthode capable de récupérer la forme 3D d'un visage à partir d'une image unique. Les entrées de cette méthode sont un modèle de visage de référence qui peut être soit un autre individu ou un visage générique et une image 2D. Les auteurs de [194] utilisent l'image d'entrée comme un guide pour « mouler » le modèle de référence afin d'atteindre la reconstruction souhaitée.

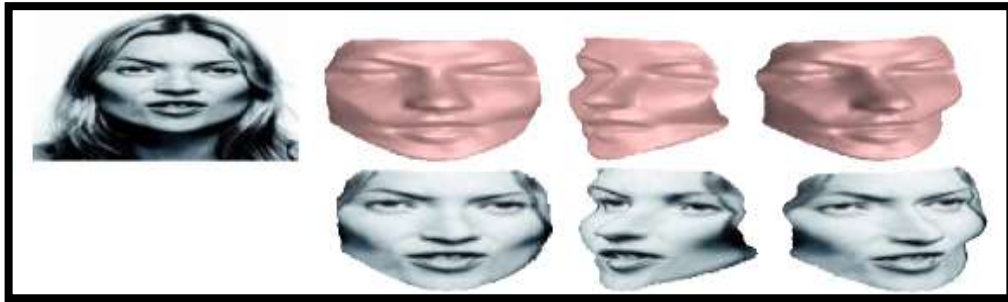


Figure 4.13. Reconstruction d'un modèle de visage 3D d'après [194].

Zollhöfer et al. [195] proposent une méthode totalement automatique pour construire un avatar personnalisé (modèle de visage 3D avec texture) avec une haute qualité à partir de la capture simultanée d'une image simple colorée et la carte de profondeur correspondante en faisant utile du capteur Kinect de Microsoft et d'un modèle 3D morphable. La méthode donnée présente à la fois l'avantage d'un enregistrement non rigoureux et d'un ajustement d'un modèle de visage morphable et fait preuve d'une grande robustesse tout en considérant qu'elle ne nécessite aucune interaction avec l'utilisateur.

Weise et al. [186] ont proposé un système qui capte et suit les dynamiques de l'expression faciale des utilisateurs en temps réel et les dessine vers un modèle virtuel (figure 4.14) en utilisant également le capteur Kinect. Le système proposé intègre le cheminement et l'animation en une seule optimisation et montre bien que la reconstruction 3D de visage et l'animation de ce dernier peuvent être réalisées en temps réel sans utilisation des marqueurs sur le visage de l'utilisateur et/ou de maquillages spécifiques, d'éclairage intrusif, ou du matériel complexe de balayage.

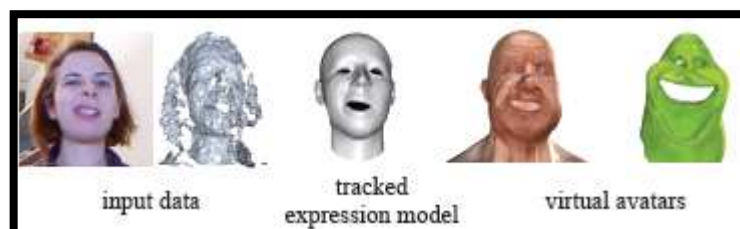


Figure 4.14. Dessin de l'expression présentée par l'utilisateur vers l'avatar d'après [186].

4.4 Deuxième contribution : La reconnaissance de visage 3D basé sur la fusion SIFT+LBP en utilisant la base de données FRGC v1.0.

4.4.1. Base de données FRGC V1.0 3D

Dans ce qui suit, nous présentons la contribution proposée pour soutenir la reconnaissance de face 3D scans en présence des parties manquantes. Pour cette expérience, nous avons utilisé l'ensemble de données FRGC v1.0 qui inclut 943 balayages de faces 3D de 275 personnes présentant une expression faciale neutre (voir Figure 4.15).

Les balayages faciaux sont donnés sous la forme d'une taille de matrice de 480x 640 de points 3D le long d'un masque de bit montrant les points disponibles du balayage (c'est-à-dire, des points proéminents correspondant typiquement aux épaules et à la tête). Étant donné que des sujets différents s'éloignent du capteur pendant l'acquisition, le nombre réel de points d'efficacité dans un balayage peut différer. Les personnes ont été acquises avec vue de face de l'épaule, avec de très petites variations dans la pose. Certaines analyses incluent des occlusions dues aux poils du visage. Plus de détails sur l'ensemble de données FRGC se trouvent dans [150]. [196]

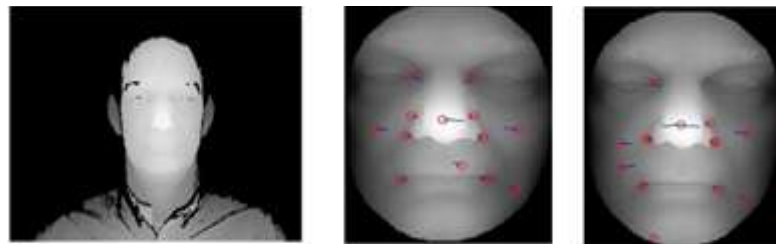


Figure 4.15. Les images de profondeur correspondent à l'image de face originale avec le jeu de données FRGCv1.0.

4.4.2 Description des étapes de l'approche proposée

Dans cette partie, nous examinons comment les caractéristiques locales de l'information 3D contribuent à la reconnaissance faciale lorsque l'éclairage, les changements d'expressions et la combinaison d'expression sous l'éclairage sont pris en compte. Tous les processus inclus dans la phase d'apprentissage et de tests sont entièrement automatiques. Notre système, comme illustré sur la figure 4.16, comprend quatre étapes importantes :

- Prétraitement : par l'alignement (translation et rotation) de l'image d'entrée 3D par rapport à l'image de référence (image de visage d'expression neutre) en utilisant l'algorithme ICP (Iterative Closest Point). Selon les images alignées, nous obtenons des images de profondeur 3D.
- Extraction des caractéristiques : une méthode d'extraction des caractéristiques robustes est très importante pour l'ensemble du système. Il est prévu que ces caractéristiques sont invariantes à la rotation, l'échelle, l'expression et l'éclairage. Dans notre système, nous utilisons les méthodes locales suivantes :
 - Transformation de caractéristiques visuelles invariantes à l'échelle (Scale-Invariant Feature Transform (SIFT)),
 - Multi-scale Local Binary Patterns (LBP).

- Classification : La méthode SVM (Support Vector Machines) est utilisée pour la réduction et la classification des données.
- Fusion : consiste en la fusion des résultats de la classification encore par la méthode SVM (Support Vector Machines) et en la normalisation de scores en utilisant la méthode Min-Max qui est choisie pour sa simplicité.

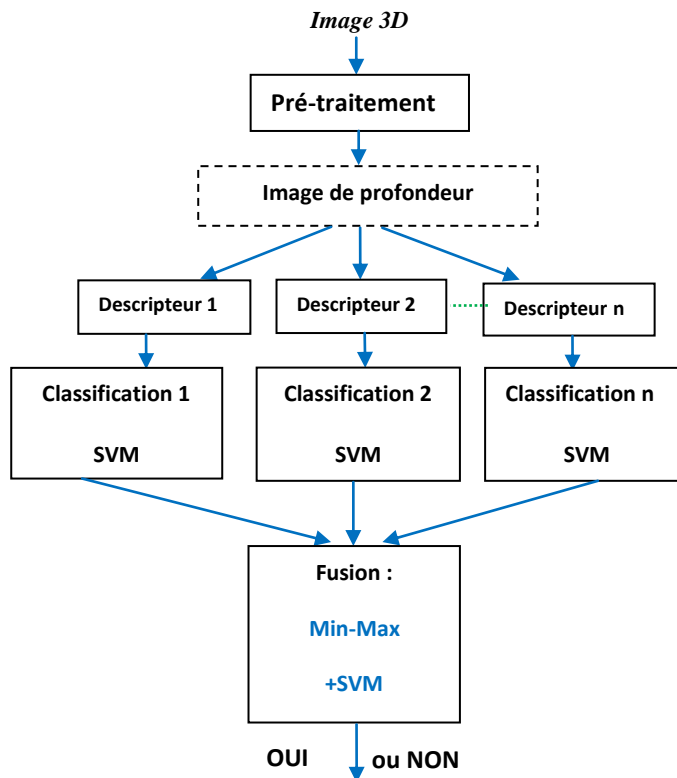


Figure 4.16. Vue de notre système proposé.

4.4.2.1 Prétraitement

On suppose qu'un visage est décrit par un nuage de points 3D capturés par un scanner laser 3D.

Chaque nuage de points est constitué de milliers de points dans l'espace 3D. Ces points discrets décrivent approximativement la surface de la face. Nous utilisons la base de données de visage FRGC v1.0. Chaque point est décrit avec des coordonnées spatiales 3D et les coordonnées de couleurs RGB correspondantes. Dans cette section, nous expliquerons comment les données 3D originales sont prétraitées. Nous alignerons les données, puis nous obtiendrons les images de profondeur et d'intensité. Cette partie prépare l'extraction de caractéristiques dans la section suivante. Ce prétraitement comporte deux étapes principales, l'alignement des surfaces de visage 3D et l'acquisition de la profondeur des images [4].

4.4.2.2 Alignement [4]

L'alignement (ou le recalage) d'entités 3D consiste à trouver un ensemble de transformations en vue de les rapprocher. Ce sujet a suscité l'intérêt des chercheurs, depuis une vingtaine d'années, compte tenu de ses nombreuses applications. Le progrès le plus significatif a eu lieu au début des années 90 avec les travaux très similaires de Besl et al. [84] et Chen et al. [197] qui décrivent deux

techniques itératives d'alignement. Nous utilisons l'algorithme du plus proche voisin itéré ICP (Iterative Closest Point) introduit par Besl et al. en 1992 dans [84]. La technique ICP permet de trouver la transformation rigide $(R; t)$ qui minimise la distance entre deux ensembles de points $P = \{p_i\}$ et $X = \{y_i\}$. Le principe consiste à déterminer, pour chaque point p_i de l'ensemble P , le point le plus proche dans le second ensemble X au sens d'une distance euclidienne. À partir de ces appariements, une transformation rigide minimisant un critère des moindres carrés (4.1) est calculée et appliquée à l'ensemble P :

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

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

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

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

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

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

4.4.2.3 Image de profondeur et d'intensité [4]

Les images de profondeur et d'intensité sont obtenues à partir des données 3D alignées. Les données sont converties en une image de profondeur 3D (voir figure 4.17 (a)) et une image en couleurs (voir figure 4.17 (b)). Dans la plupart des images, le nez est la partie du visage la plus proche au scanner 3D. Pour chaque pixel, la moyenne est calculée en utilisant une fenêtre de taille 9×9 autour de lui.

L'utilisation d'une fenêtre de taille 3×3 qui calcule la somme de la profondeur de ses pixels correspondants, le nez est détecté en tant que coordonnées du pixel de la fenêtre centrale qui retourne la valeur minimale. Après détection du nez, on extrait une sous-image centrée sur le centre de nez. Toutefois, en raison de la qualité originale des données 3D, les images de profondeur et d'intensité contiennent habituellement beaucoup de bruit, tels que des trous et des valeurs aberrantes. Nous pouvons obtenir des images améliorées par les procédés suivants :

Pour chaque pixel, la moyenne est calculée pour la fenêtre de 5×5 . Si la valeur du pixel est inférieure à un seuil donné, ce pixel est remplacé par le pixel moyen. L'égalisation d'histogramme est utilisée pour réduire l'influence des variations d'illumination.



Figure 4.17. Exemple de l'image de profondeur et d'intensité.

4.4.2.4 Extraction des caractéristiques du visage [4]

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

- Scale-Invariant Feature Transform (SIFT) se compose de caractéristiques invariantes à l'échelle et à la rotation.
- Local Binary Patterns (LBP) : Récemment, la représentation LBP a gagné beaucoup d'attention dans le domaine de la reconnaissance faciale et est également appliquée avec succès en reconnaissance faciale 2D. Inspirés par l'opérateur LBP 2D, nous avons développé un algorithme de reconnaissance de visage 3D.
- La fusion des deux méthodes (SIFT+LBP).

Le but principal de ces méthodes est de trouver une base de projection qui optimise un certain critère. Cette base est ensuite projetée sur une base de dimension inférieure. Dans notre travail, nous avons testé ces méthodes sur la base de données FRGC v1.0 du visage 3D. Nous avons montré que la méthode SIFT+LBP est meilleure que d'autres méthodes d'extraction de caractéristiques.

Les paragraphes suivants présentent le calcul de ces caractéristiques.

➤ *Scale-Invariant Feature Transform (SIFT)*

Scale-Invariant Feature Transform (SIFT), que l'on peut traduire par « transformation de caractéristiques visuelles invariantes à l'échelle », est une approche pour la détection et l'extraction de descripteurs de caractéristiques locales qui sont raisonnablement invariantes aux changements dans l'éclairage, le bruit d'image, la rotation, la mise à l'échelle et de petits changements de points de vue. Il a été développé en 1999 par le chercheur David Lowe [56]. Selon [56], la base de l'algorithme SIFT se compose de cinq étapes : [4]

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

Deux points clés SIFT détectés sur une image de plage pour identifier un profil de visage, sa fonction 1D correspondant aux valeurs de profondeur correspondant aux points du segment qui relie les deux points de touche. Plus formellement, soit :

- $I(x)$ où $x \in \mathbb{R}^2$ est l'image d'une plage de balayage du visage ; x_1, x_2 deux points-clés ;

Alors, le profil identifié par la paire ordonnée (x_1, x_2) est défini comme :

$$P_{i,x_1,x_2}(t) = I((1-t)x_1 + tx_2) \quad t \in [0,1] \quad (3)$$

En décrivant la face de modèle proposée, des traits distinctifs du visage sont capturés en retenant les descripteurs de points clés SIFT détectés sur l'image de plage et les profils identifiés par des paires de ces points clés. Les données de résultat organisées dans une structure de graphe avec des nœuds corrélés aux points de clés et aux arêtes corrélées aux profils.

Étant donné les graphiques des deux faces, la dissimilarité est mesurée par le premier attribut à chaque nœud du graphe son premier nœud le plus proche présent dans le dernier graphe, la proximité est calculée comme la distance euclidienne entre les descripteurs SIFT 128 dimensions liés aux points clés. [198] Ensuite, pour chaque paire de nœuds correspondants dans les deux graphes, des profils entiers construits à partir des points de clé sont différenciés pour associer les deux profils à la distance minimale. Déterminer le profil générique P1 extrait du couple de points clés (x_1, x_2) de l'analyse de visage I1 et du profil P2 extraire du couple de points clés (x_3, x_4) de l'analyse de visage I2 la distance entre les deux profils est mesurée :

$$D(P1(t), P2(t)) = \int_0^1 |P1(t) - P2(t)| dt \quad (4)$$

Chapitre 4: Notre méthode pour la vérification des visages 3D

Enfin, la distance entre les faces des deux balayages est mesurée par les profils de valeurs de distance minimale moyenne sur chaque paire de nœuds obturés. Nous cherchons le classificateur linéaire qui sépare les données avec l'erreur de génération la plus faible. Ce classificateur est un hyperplan maximisant la marge d'erreur, qui est la somme des distances entre l'hyperplan et les exemples (positif, négatif) de l'hyperplan le plus proche [199] ; exige la classification de ces deux classes appliquées la méthode SVM qui sera présentée à l'étape suivante d'apprentissage. Dans le cas où les données ne peuvent pas être séparées par une fonction linéaire, une non-linéarité peut être introduite en utilisant une fonction non linéaire. Enfin, la distance entre les deux balayages de faces est calculée en faisant la moyenne des valeurs de distance minimale de profil au-delà de toutes les paires de nœuds correspondants.

Par conséquent, le schéma peut être utilisé sans aucune modification pour supporter l'appariement d'un simple balayage partiel d'un visage contre un balayage complet du visage. Pour cette raison, la solution proposée permet la reconnaissance faciale même en présence de pièces manquantes et/ou d'occlusions.

La représentation du modèle 3D concernait la détection de visage sur la base d'un certain nombre de points clés sur la surface de face 3D et la description de la surface de la face 3D correspondant aux points de clé et le long des chemins de liaison linéaire de paires de points de clé. Contrairement aux possibilités que les points clés doivent correspondre face à des repères importants, tels que les sourcils, les yeux, la bouche, les joues et le nez, nous nous en tenons à une supposition particulière sur la position attendue de la clé de points sur la surface du visage. Relativement, nous incluons la position des points clés pour influencer par des traits morphologiques spécifiques de chaque sujet. [198]

Dans notre contribution, nous avons utilisé l'algorithme SIFT pour identifier les points clés et la description. Les techniques SIFT ne sont adaptées pour les images 2D en niveaux de gris et ne peuvent pas être appliquées immédiatement aux balayages de visage 3D. Bien que les informations 3D des faces numérisées puissent être capturées par des images de plage qui utilisent l'échelle de gris de chaque pixel pour représenter la profondeur d'un balayage de visage. Bien qu'il soit possible de détecter un grand nombre de points clés à différents emplacements dans une image, seuls les invariants les plus distincts et les plus utiles pour la compatibilité doivent être conservés. Ceux-ci tombent fréquemment sur les bords et les coins de l'image, et peuvent être de beaucoup de différentes tailles et orientations aussi bien. Pour détecter les points clés, les descripteurs SIFT sont calculés. En bref, un descripteur SIFT pour une petite région de l'image, par ex. La taille 4x4 est calculée à partir des histogrammes de gradient de vecteur des pixels dans le patch. Huit directions de gradient probables et donc la taille totale du descripteur SIFT est donnée par l'expression $4 \times 4 \times 8 = 128$ éléments. Ce descripteur est unifié pour améliorer l'invariance aux changements d'éclairage (non pertinent dans le cas de la plage d'images), et effectué par d'autres moyens pour assurer l'invariance de l'échelle et la rotation ainsi. Ces propriétés font que le descripteur SIFT peut émettre un solide et puissant représentant local de la portée de l'image et, par conséquent, de la surface du visage.

En particulier, les descripteurs SIFT ont été extraits en utilisant le paramètre suivant :

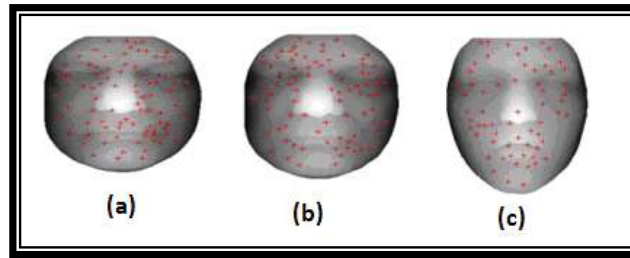


Figure 4.18. Gamme d'images à partir de trois balayages de visage 3D montrant les points SIFT détectés avec des informations de profondeur 3D.

- Les points clés SIFT sont extraits et conservent leurs angles de portée et d'orientation ; - 4x4 Histogrammes d'orientation les régions de chaque point clé quatre échantillonnages sont utilisés pour calculer le descripteur SIFT [198].

Comme dans la figure 4.18, les points clés détectés par SIFT sont présentés pour la gamme d'images de trois différentes analyses du visage. Les deux premières images représentent le même sujet, tandis que le troisième est un sujet différent. Dans la largeur, la position des points clés sur un scan de visage détecté dépend des caractéristiques morphologiques du visage. Ceci est confirmé par le fait que l'arrangement spatial des points clés détectés sur des scans différents d'un même sujet | Le long d'une expression neutre | Est très comparable. Cependant, l'information qui est captée en combinant des descripteurs locaux (par exemple des descripteurs SIFT) avec des points clés détectés dans une analyse faciale n'est pas suffisamment discriminante pour permettre une reconnaissance précise de l'identité de l'objet. Cela s'explique par plusieurs raisons, qui se rapportent le plus significativement à la notoriété des descripteurs locaux et inter-sujets itérés. Ce dernier fait référence à l'absence de toute garantie que la position des points clés sur le modèle de face doit être séparée sur le sujet. En effet, puisque la position des points clés dépend de la morphologie du visage, la détection de points clés dans des modèles de sujets différents comprend une disposition différente sur le visage. Autour de la salience des descripteurs locaux, calculés à des points clés détectés, il convient d'observer que ces descripteurs captent la morphologie de la surface de l'image dans un petit voisinage (généralement 16x16 pixels) centré au point clé. Le rapport qui est retenu est juste assez pour permettre la reconnaissance de la partie du visage qui est le point clé : le nez, l'un des points extrêmes de la bouche ou des yeux, etc. [198].

Toutefois, cette information n'est pas suffisante pour souligner la reconnaissance d'une personne spécifique.

➤ **Multi-scale local binary patterns (LBP)**

Compte tenu de l'image de profondeur et d'intensité du visage, nous générons un ensemble de multi-échelles LBP pour la représentation du visage. Quelques exemples sont illustrés dans la figure 4.19. Sur cette figure, le nombre de points d'échantillonnage varie de 8 points à 24 points et la valeur du rayon varie de 1 pixel à quatre pixels.

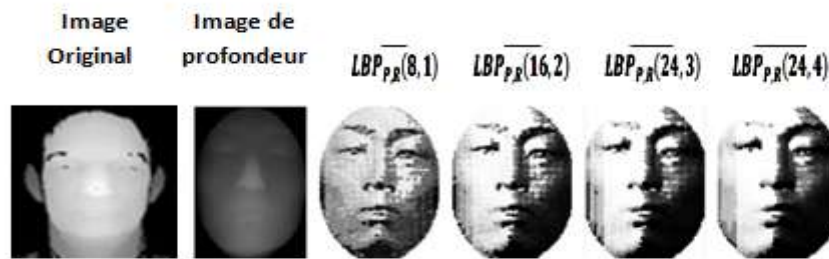


Figure 4.19. Multi-échelles LBP de l'image de profondeur et d'intensité du visage [4].

➤ **La fusion LBP+SIFT**

Nous avons appliqué l'opérateur SIFT sur les images $LBP_{p,R}$ séparément. Parce que la $LBP_{p,R}$ met en évidence les caractéristiques locales lisses de l'image de profondeur. On peut mieux détecter un nombre de points-clés par l'utilisation de SIFT sur l'image $LBP_{p,R}$ que l'utilisation de SIFT dans les images originales. Nous avons effectué un travail statistique sur la base de données FRGC v1.0. Le nombre moyen de points-clés extraits de chaque image $LBP_{p,R}$, est égal à 52 pour les images de profondeur et 162 pour les images d'intensité. En revanche, le nombre moyen de points-clés extraits de chaque image faciale d'origine de profondeur est limité à 14. Pour l'intensité, le nombre de points-clés est limité à 63. Nous avons utilisé le code Lowes [200] pour calculer le descripteur [4].

La figure 4.20 montre les points-clés extraits à partir de l'image du visage originale et de ses quatre images $LBP_{p,R}$ associés.

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

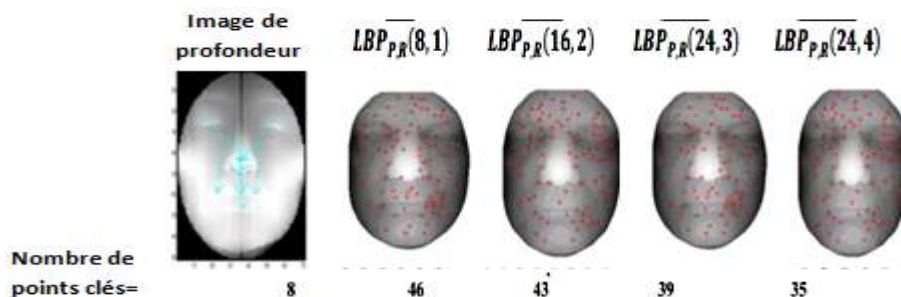


Figure 4.20: Points-clés détectés avec le descripteur SIFT de l'image de profondeur et d'intensité originale et les quatre images $LBP_{p,R}$ associées.

4.4.2.5 Classification [4]

La classification est utilisée par l'SVM pour minimiser l'erreur de sur l'ensemble d'apprentissage alors la marge est la distance quadratique moyenne entre la séparatrice et les éléments d'apprentissage les plus proches de celle-ci appelés vecteurs supports (voir chapitre 3).

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

Chapitre 4: Notre méthode pour la vérification des visages 3D

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

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^M \alpha_i (y_i (w \cdot K(x_i) + b) - 1) \quad (5)$$

Avec les coefficients α_i appelés multiplicateurs de Lagrange.

Pour résoudre ce problème d'optimisation, il faut minimiser le Lagrangien par rapport à w et b et le maximiser par rapport à α . Dans la pratique, il est souvent impossible de trouver un séparateur linéaire (même dans l'espace transformé par la fonction noyau) car il y a toujours des erreurs de classification. Vapnik [149] a donc introduit la technique de "marge souple". Ce principe de marge souple tolère les mauvais classements par l'introduction de variables ressorts ζ_i qui permettent de relâcher les contraintes sur les éléments d'apprentissage qui ne doivent plus être à une distance supérieure ou égale à 1 de la marge (l'égalité correspondant aux vecteurs de support), mais à une distance supérieure ou égale à $1 - \zeta_i$, c'est à dire :

$$y_i (w \cdot K(x_i) + b) \geq 1 - \zeta_i \quad (6)$$

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

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

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - C \sum_{i=1}^M \zeta_i - \sum_{i=1}^M \alpha_i (y_i (w \cdot k(x_i) + b) - 1) \quad (7)$$

Où C est une constante strictement positive qui permet de régler le compromis entre le nombre d'erreurs de classification et la largeur de la marge [150].

4.4.3 Scores Fusion

Un système de fusion est constitué de deux modules, un module de fusion et un module de décision (voir figure 4.21). Le problème devient donc un problème de classification à 2 classes (OUI ou NON, Client ou Imposteur) à partir d'un vecteur de nombre réels dont la dimension est égale au nombre de sous-systèmes [4].

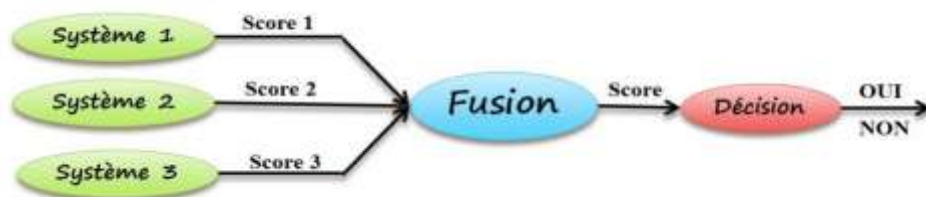


Figure 4.21. Schéma de la fusion des scores [4].

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

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

Nous avons étudié deux approches pour la fusion de scores et plusieurs méthodes de normalisation des scores (voir chapitre 3), mais dans ce chapitre et en particulier cette contribution, nous avons utilisé comme normalisation de scores la méthode Min-Max (équation 8) pour sa simplification.

- Normalisation par la méthode Min-Max [4] :

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

$$S'_{ik} = \frac{s_{ik} - \min(\{s_i\})}{\max(\{s_i\}) - \min(\{s_i\})} \quad (8)$$

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

Pour la fusion, nous avons utilisé une méthode de classification qui est Support Vector Machines (SVM), parce qu'il s'agit d'une méthode efficace et bien connue dans le domaine de la classification en deux classes. Nous utilisons une machine vectorielle de support (SVM) avec noyau RBF (fonction de base radiale) (voir équation 9). SVM a été mis en œuvre en utilisant le LibSVM qui est une bibliothèque disponible gratuitement [201]. Le noyau RBF utilisé est de la forme :

$$K_{\text{RBF}}(u, v) = e^{-\gamma \|u-v\|^2} \quad (9)$$

Où γ est un paramètre qui ajuste la largeur de la marge. SVM nécessite également le paramétrage du paramètre C qui ajuste l'équilibre entre la largeur de la marge et le nombre d'erreurs admissibles (comme décrit dans l'équation 9). Les différents taux d'erreur à l'aide de la fusion par une machine de vecteur support (SVM) avec noyau RBF sans et avec des méthodes de normalisation [4].

4.5. Conclusion

Dans ce chapitre, nous avons détaillé, dans la première partie, notre solution pour l'identification des expressions faciales en utilisant la caméra Kinect. Pour se faire, nous avons donné le schéma bloc de l'approche ainsi un schéma descriptif des grands titres pour chacune de ses étapes. Dans la deuxième section, nous avons présenté le capteur Kinect que nous proposons d'utiliser pour l'acquisition de données. Dans les deux sections qui suivent, nous avons expliqué les notions théoriques et les outils adoptés respectivement pour les étapes détection et reconnaissance, à savoir la technique « Viola&Jones » et la réduction d'espace par la méthode « MB-LBP ».

Dans la deuxième partie, nous avons présenté notre contribution pour la reconnaissance de visage 3D en utilisant multi-algorithmes de visages qui est testé sur une base de données contrôlée FRGC v1.0. Nous avons utilisé des approches locales basées sur la réduction d'espace de données c'est la fusion des deux méthodes « SIFT+LBP », car elle reste parmi les techniques les plus utilisées et les plus populaires dans le domaine de la reconnaissance des visages, et la classification de ces données est utilisée par la méthode SVM qui a la particularité d'être plus simple à mettre en œuvre, l'SVM agit directement sur la précision en minimisant l'erreur de généralisation et donne de bons résultats. Dans le chapitre suivant, nous allons présenter les étapes d'implémentation des deux contributions proposées ainsi que les expérimentations effectuées.

Implémentation, résultats et comparaison

5.1. Introduction

Depuis l'avènement des technologies permettant l'affichage et la manipulation d'objets et de scènes tridimensionnelles, les travaux de recherche se sont multipliés pour viser et solutionner des problèmes appartenant à différents domaines d'application. Nous pouvons citer le domaine médical, la sécurité et le contrôle, le prototypage rapide, la biométrie de visages, les jeux vidéo et l'animation.

Dans ce chapitre, nous consacrons la première section pour présenter notre solution visant à exploiter l'expression faciale et l'état émotionnel de l'utilisateur pour la conception d'une interface permettant une compréhension des mouvements du visage d'un ensemble de personnes (10 personnes à raison de 10 images/personne) et de réagir en conséquence. En concordance avec la position du visage par rapport à la caméra (en avant ou en arrière), notre application permet de zoomer ou de dézoomer. La migration du visage de l'utilisateur d'une expression à une autre sera aussi affichée par un modèle tridimensionnel. Nous présentons également les outils que nous avons exploités dans l'implémentation de notre solution. Nous décrivons, par ailleurs, les activités de notre système, et présentons les interfaces graphiques de notre application qui a été développée en langage C++. Enfin, nous effectuerons des comparaisons de nos résultats avec ceux de la littérature scientifique.

Dans la deuxième section, nous présentons les résultats de l'utilisation de la représentation et de l'appariement pour assurer la reconnaissance de visages tridimensionnels scannés présentant la particularité d'avoir des parties manquantes. Pour cette expérience, nous avons utilisé l'ensemble de données FRGC v1.0, qui inclut 943 balayages de visages 3D de 275 personnes présentant une expression faciale neutre. Nous présentons les outils que nous avons utilisés dans l'implémentation de notre solution développée en langage C++. ainsi que les résultats obtenus. Enfin, les résultats sont comparés avec ceux de la littérature scientifique.

5.2 Contribution I : Implémentations, résultats et comparaison

5.2.1 Système automatique "Temps réel" :

Notre solution cherche à exploiter l'expression faciale et l'état émotionnel de l'utilisateur pour la conception d'une interface permettant une compréhension des mouvements du visage et de réagir en conséquence. En concordance avec l'avancée et le recul du visage par rapport à la caméra,

notre application permet de zoomer et/ou de dézoomer. La migration du visage de l'utilisateur d'une expression à une autre sera aussi affichée par un modèle de visage 3D.

Dans cette section, nous consacrons la première section à la présentation des outils que nous avons exploités dans l'implémentation de notre système. Une deuxième section est dévolue à la description du fonctionnement de notre système que nous avons nommé **AREFK** (**A**utomatic **R**ecognition **E**xpression of **F**ace with **K**inect). La troisième section sera réservée à la présentation des interfaces graphiques de notre application développée en langage C++.

5.2.1.1. Outils utilisés :

❖ Dispositifs matériels

Notre configuration matérielle inclut les dispositifs suivants :

- Un PC avec Windows 7 et 4 Go de RAM
- Capteur Webcam.

❖ Outils logiciels

Le langage C++ : Nous avons choisi d'utiliser C++ comme langage de programmation pour développer notre système, la librairie OpenNI « Open Natural Interaction » : C'est une bibliothèque qui permet la reconnaissance des gestes de l'utilisateur et de les interpréter comme étant des commandes. Elle est destinée à agir comme interface entre une variété de pilotes de capteurs (le capteur Webcam dans notre cas) et le middleware pour manipuler les données capturées à travers des commandes de programmation ayant un niveau d'abstraction plus haut et donc simple à utiliser.

La librairie OpenCV « Open Source Computer Vision Library » : C'est une bibliothèque qui offre des fonctions de traitement d'image par ordinateur. Toutes ses fonctionnalités ont été programmées en langage C/C++.

5.2.1.2. Description des activités de notre système

❖ Méthodologie de détection

L'étape de détection permet de décider exclusivement sur l'existence ou non d'un visage dans une image. Nous avons donc besoin d'un détecteur (classifieur) de visages permettant de chercher les caractéristiques relatives à ce dernier.

La bibliothèque OpenCV implémente la méthode de détection de visages « Viola&Jones » et offre comme résultat une liste de fichiers « .XML » dits classifieurs en cascade de Haar. Nous avons donc pu éviter la phase d'apprentissage d'un détecteur en exploitant le classifieur évalué dans le fichier « haarcascade_frontalface_alt2.xml » destiné à détecter un visage en particulier lorsqu'il est présent en vue frontale.

Nous avons aussi utilisé les fonctions de cette même bibliothèque pour détecter le visage dans chaque frame de la séquence vidéo. Les coordonnées du visage détecté sont passées à une fonction chargée d'englober ce dernier dans un rectangle tout en traitant les séquences vidéo image par image.

Notre détection a pu suivre implicitement le mouvement du visage détecté. Nous avons aussi synchronisé le dessin de rectangles englobants avec la détection afin qu'ils prennent en charge toute modification de position et de dimension du visage (figure 5.1).



Figure 5.1. Détection de visage et dessin de rectangle englobant dans chaque trame en utilisant « haarcascade_frontalface_al2.xml ».

❖ Définition d'une segmentation spatiale du visage.

Après avoir détecté le visage, nous avons utilisé l'approche géométrique pour localiser les régions contenant les éléments significatifs du visage tels que la bouche, les yeux, le nez... Nous entendons par le terme « approche géométrique », la méthode qui exploite les connaissances a priori relatives à la répartition spatiale du visage humain. Autrement dit, la géométrie d'un visage humain est toujours formée de haut vers le bas d'un front, sourcils, yeux, nez, bouche et menton. Ainsi, au niveau du nez et dans le sens horizontal, elle présente l'oreille gauche, joue gauche, nez, joue droite et oreille droite. En nous basant sur cette répartition, nous avons proposé une segmentation spatiale du visage dont le résultat est donné par la figure 5.2.

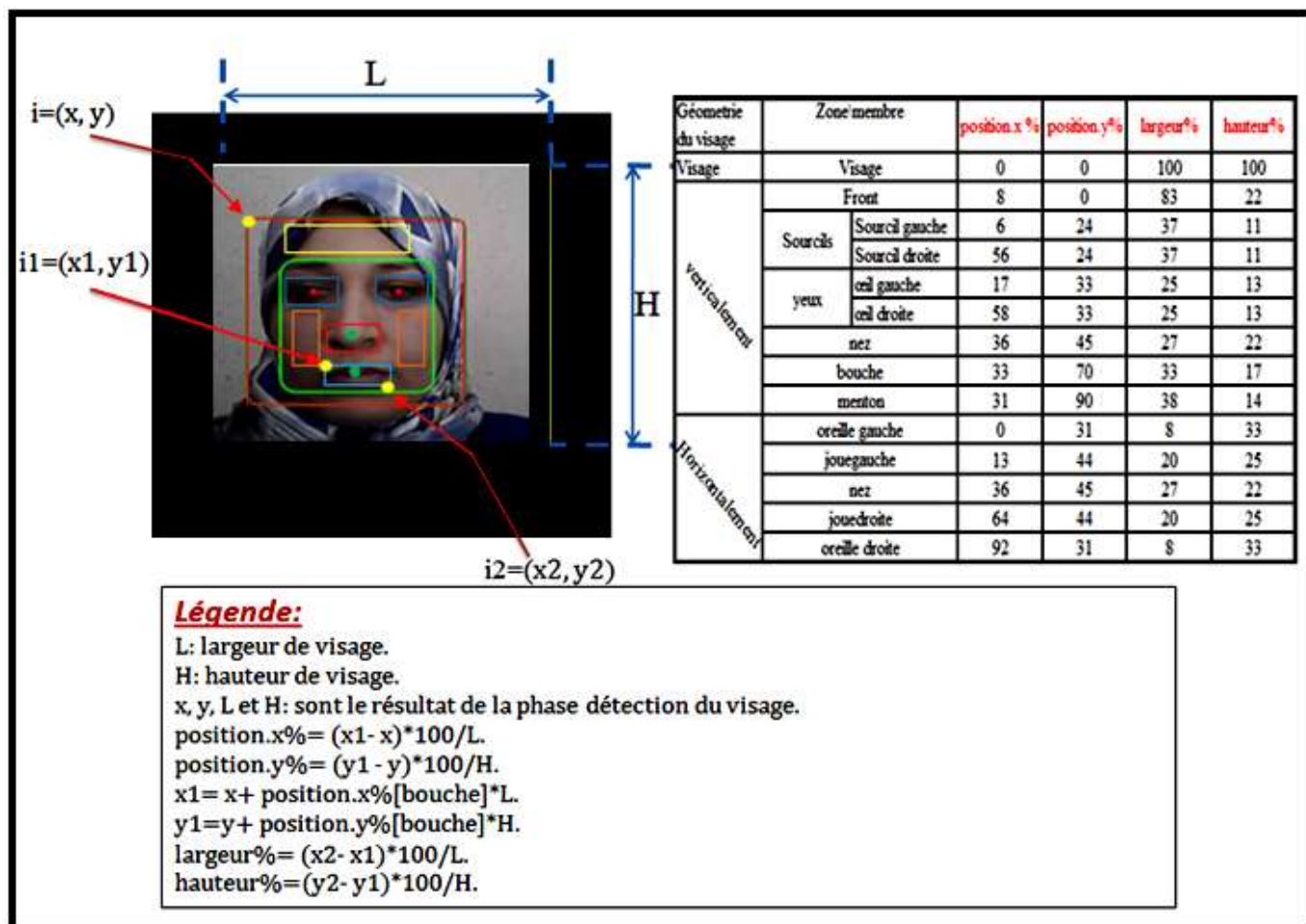


Figure 5.2. Segmentation spatiale du visage.

Dans la figure 5.2, nous expliquons comment nous avons traité la détection des yeux selon le visage avant de déterminer les coordonnées du rectangle englobant la bouche et ce, après avoir récupéré les coordonnées du visage. La phase de détection fournit, comme résultat, les coordonnées du point initial du visage (c'est le point $i = (x, y)$), la largeur L et la hauteur H du visage. Par la suite, ces informations sont transmises comme paramètres à une fonction qui dessine un rectangle marquant le visage détecté. Cette dernière calcule le point initial de la bouche (dans notre cas c'est le point $i_1 = (x_1, y_1)$), la largeur et la hauteur de la bouche à partir des données reçues et dessine également le rectangle résultat. Le calcul des coordonnées est décrit en légende dans la figure 5.2. Nous localisons le reste des différents traits faciaux comme l'indique le tableau à droite de la figure 5.2.

À travers cette localisation, nous avons la possibilité d'enregistrer les différentes zones du visage, chacune à part, sous la forme d'une image comme illustré dans la figure 5.3 ci-dessous.



Figure 5.3. Membres de visage localisés et enregistrés dans des images.

❖ Notre solution pour la réalisation du calibrage

À notre connaissance, il existe deux méthodes de calibrage : une méthode de calibrage dite passive et une seconde dite active. La première, la plus couramment utilisée, consiste à employer soit deux caméras, soit une seule avec acquisition de deux positions différentes. Dans notre cas, nous nous intéressons à la deuxième technique qui utilise une seule caméra puisque nous utilisons les flux vidéo d'une caméra.

Afin de déterminer les paramètres des caméras (RGB et IR) de notre dispositif d'acquisition, nous utilisons un code source développé par Nicolas Burrus qui fait recours à certaines fonctions de la bibliothèque OpenCV et fournit en résultat six fichiers « .XML » représentant la matrice K, les matrices de disparité de la caméra, la matrice de rotation et le vecteur de position de la caméra. Le programme nécessite la capture de 80 images couleur d'un échiquier dont la structure réelle est connue (nous utilisons le damier d'OpenCV2.3) et 80 images de pattern IR. Nous désignons par structure, le nombre de carreaux noirs horizontaux, verticaux ainsi que la largeur réelle des carreaux. Il est essentiel de noter que nous redéfinissons les valeurs de ces critères puisque nous utilisons un échiquier différent de celui utilisé par Nicolas. D'un autre côté, puisqu'il n'est pas possible d'activer simultanément les modes IR et RGB de la caméra, nous avons besoin donc d'activer ces deux modes de manières alternatives.

Pour calibrer la caméra, nous avons élaboré un code C++ permettant d'alterner entre ces deux modes pour collecter les images de la mire de calibrage demandées par le projet C++ déjà cité. Le pseudo-algorithme ci-dessous décrit notre procédure de capture. Nous avons enregistré les images dans des répertoires sous les noms view00ij avec i variant de 0 à 7 (représentant le numéro de l'itération -1) et j variant de 0 à 9 (représente le numéro de l'image dans chaque itération). Chaque répertoire se compose de deux fichiers images color.bmp (l'image couleur) et intensity.bmp (l'image IR). Ces contraintes ont été fixées par le code source utilisé.

1. **Positionnement fixe de caméra ;**
 2. **Pour i de 1 à 8 faire**
 - **Fixation (i) de la mire de calibrage ;**
 - **Lecture d'un caractère à partir du clavier pour lancer l'itération ;**
 - **Activation du mode RGB ;**
 - **Capture et sauvegarde de 10 images RGB ;**
 - **Activation du mode IR ;**
 - **Capture et sauvegarde de 10 images IR ;**
 - **Application du filtre gaussien sur les images IR afin de faciliter la détection des coins du damier ;**
 - **Affichage de message « Fixer de nouveau votre échiquier » ;**
- Fin Pour**

Figure 5.4. Pseudo-algorithme de la procédure de collecte.

Après avoir rassemblé 160 images réparties sur 80 dossiers dont chacun possède une image couleur et une image IR. Nous avons accédé alors au niveau code source du code C++ pour spécifier le chemin du dossier parent des images collectées puis nous démarrons l'exécution. Pour chaque paire d'images, le programme cherche les différents coins d'intersection du damier et établit la correspondance entre eux dans les deux images. Il calcule les paramètres intrinsèques de la caméra ainsi que ses paramètres extrinsèques (matrice R et vecteur t) ainsi que les deux matrices de disparité relatives à la caméra. L'erreur de la caméra est visualisée par la suite. Le résultat de ce dernier calcul est la génération d'un fichier « .XML » pour chaque paramètre déterminé.

Dans le but d'aligner les flux de caméra, nous avons pu aussi tester la fonction `SetViewPoint(image)` de la classe `DepthGenerator` (classe responsable de la génération du flux de profondeur) offerte par la librairie `OpenNI` et servant à définir le point de vue de `DepthGenerator` pour le correspondre à celui de `ImageGenerator` (classe responsable à générer le flux image RGB). Une fois le visage détecté, l'alignement des flux permet de localiser l'information pertinente relative au visage et de translater les coordonnées du rectangle englobant ce dernier.

Dans la figure 5.5, les deux premières images (a) montrent clairement la subdivision de la carte des disparités sur l'utilisateur au centre. Les deux autres images à droite présentent une carte des disparités avec moins de dédoublement. Cela montre l'importance de calibrer la caméra avant toute récupération de données.



Figure 5.5. Alignement des deux flux caméra : (a) visualise les deux flux avant l'appel de la fonction `SetViewPoint()` et (b) visualise l'effet de la fonction.

❖ Notre étape de Prétraitement

Après la détection du visage sur les deux flux, l'étape suivante consiste à éliminer les informations de l'arrière-plan.

Nous procédons à l'extraction du visage en deux phases. Au cours de la première, nous éliminons tout ce qui est en dehors du rectangle englobant le visage détecté dans les deux flux. Dans la deuxième, nous nous limitons au flux de profondeur sur lequel nous effectuons un seuillage pour enlever l'information supplémentaire à l'intérieur du rectangle dessiné. Nous utilisons la fonction `cvThreshold()` offerte par OpenCV avec comme valeur de seuil 200 et un type de seuillage `CV_THRESH_TOZERO` qui permet de sauvegarder tous les pixels de l'image qui sont supérieurs au seuil indiqué et de rendre le reste à zéro. La figure 5.6 montre le résultat des deux phases effectuées.



Figure 5.6. Extraction de visages des deux flux.

❖ Notre solution pour la reconnaissance

Le processus de reconnaissance s'appuie sur les deux phases précédentes de caractérisation et d'**apprentissage**. Il implique plusieurs images de visages ayant des expressions. Pour identifier l'expression de la personne présente en face de la caméra, nous avons d'abord testé la base d'images offerte par Andriew Davision. Vu que la luminosité des images issues de la caméra RGB est différente de celle de la petite base testée, les résultats de reconnaissance ont été faibles.

Nous avons alors choisi de créer notre propre base d'expressions pour échapper à la limite de classification par « Eigenfaces » par rapport au SVM, lors d'une variabilité interclasses associée à une large variabilité de lumière au sein d'une même classe. La figure 5.7 ci-dessous montre des exemples des expressions faciales de notre base.



Figure 5.7. Quelques images d'expressions de notre base d'apprentissages.

L'étape de caractérisation crée une liste d'images dites « Eigenfaces » dont le nombre est égale au nombre d'images de la base. Dans notre cas, nous appelons une image de sortie de cette étape une « EigenExpression » vue quelle présente les éléments pertinents qui permettent de distinguer entre les différentes expressions de notre base. La figure 5.8 ci-dessous montre certaines EigenExpressions résultantes de cette étape.

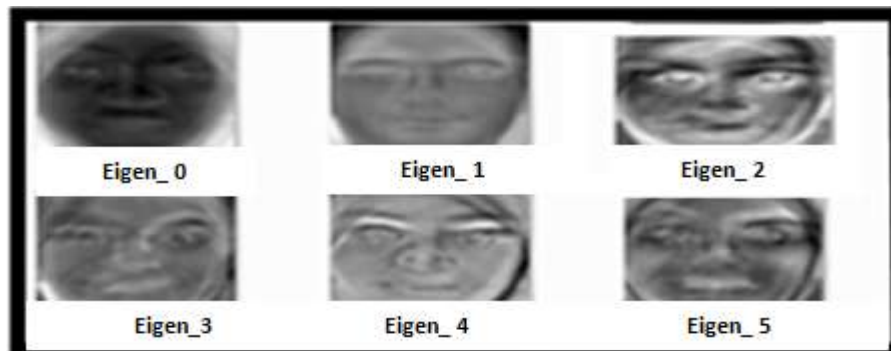


Figure 5.8. Certaines EigenExpressions calculées à partir de notre base.

Une fois les EigenExpressions construites, nous déterminerons la série des poids de chaque image de notre base. Autrement dit, nous calculons les coordonnées de chaque expression dans l'espace engendré par les EigenExpressions élaborées.

La réduction de l'espace présente un cas de projection de l'expression par exemple « Joie » dans un espace réduit des EigenExpressions calculées par la méthode MB_LBP. Ceci dans le but de calculer le vecteur de poids de chaque expression dans l'espace engendré par MB_LBP pour toutes les EigenExpressions.

L'apprentissage sert à maintenir les informations utiles à la classification. Dans le cas de la méthode « EigenFace », cette phase sauvegarde seulement les X premiers vecteurs de poids sous forme d'une matrice dite « matrice de poids ». L'apprentissage se fait donc par le SVM, il est essentiel de noter que X choisit pour notre implémentation est égale à M (nombre d'images de notre base).

Nous traitons désormais la reconnaissance proprement dite. Au cours de cette étape, nous réalisons l'extraction du visage de l'utilisateur, ensuite nous calculons le vecteur de poids de l'expression présentée par le visage capturé et enfin nous comparons le vecteur résultat avec chaque colonne de la matrice de poids. Nous déterminons la classe de l'expression de test en nous basant sur la distance euclidienne minimale. Ce choix nous semble bon pour équilibrer le rapport performance et complexité du système.

❖ Génération du fichier « .pcd » :

Le capteur de profondeur génère une grille de 640×480 qui offre la distance de chaque pixel vis-à-vis le Kinect. Cette distance présente la coordonnée Z du point réel. Comme illustré par la figure 5.9, la connaissance de la coordonnée z du point permet de calculer son ordonnée. Un autre calcul permet d'obtenir son abscisse [131].

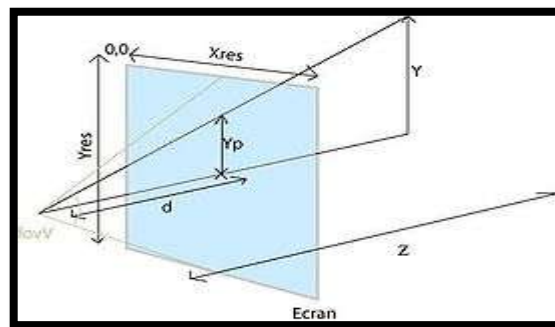


Figure 5.9. Calcul de l'ordonnée y d'un point à partir de z [131].

Le traitement de toute la carte générée par Kinect donne un nuage de points 3D représentant la scène capturée dont les dimensions sont les mêmes que les dimensions réelles. Nous calculons les coordonnées de chaque pixel de la grille produite par le Kinect et nous sauvegardons le résultat dans un fichier « .pcd ».

❖ Visualisation des modèles 3D du visage

Une fois le visage détecté et l'expression manifestée par l'utilisateur reconnue, la tâche qui reste à faire par notre système, c'est l'affichage de modèle 3D du visage représentant cette expression. Afin d'atteindre notre objectif, nous nous servons du logiciel FaceGen Modeller pour générer sept modèles du visage 3D (figure 5.10 montre des exemples des sept modèles du visage 3D) sous format de fichiers « .obj » dont chacun présente une expression parmi celles étudiées, se résume à l'importation de sept modèles et l'affichage du modèle possédant l'expression visualisée. Autrement dit, nous aurons un changement de forme du modèle visualisé vers la forme du modèle représentant l'expression identifiée. Il est indispensable de mentionner que nous nous servons des fonctionnalités offertes par l'API « JAVA 3D » [131].

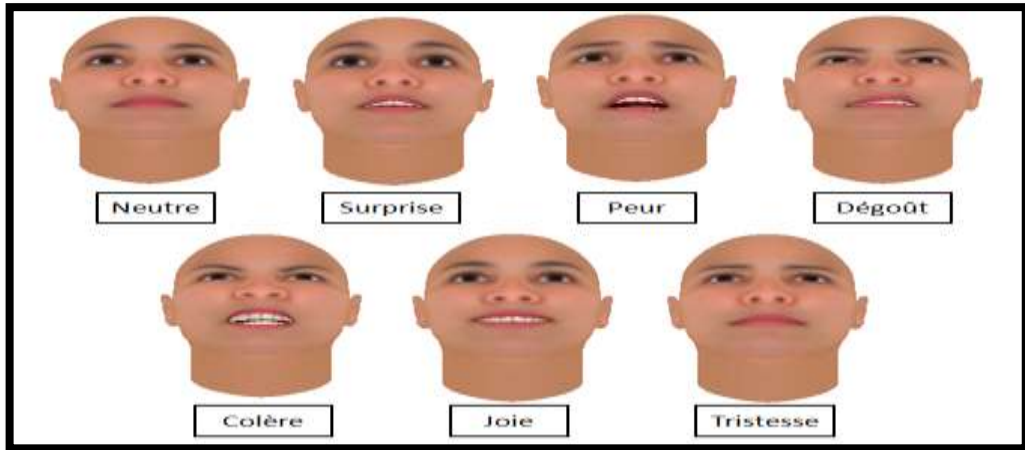


Figure 5.10. Sept modèles de visage 3D génériques [131].

5.2.2 Résultats : Présentation des interfaces de notre Système

La première interface affichée lors de lancement de notre application est celle présentée par la figure 5.11 qui sert à visualiser le flux capturé par la caméra.

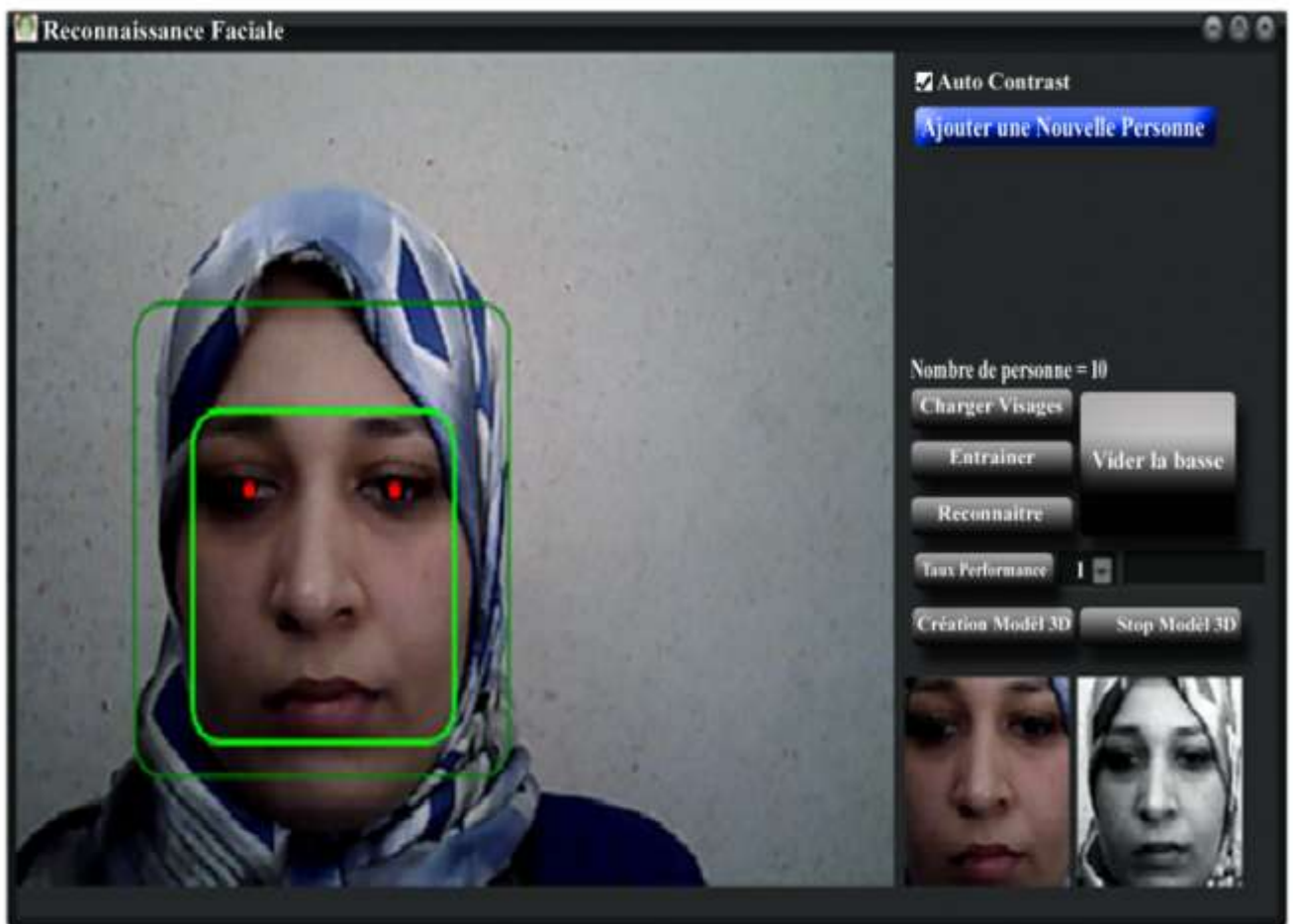


Figure 5.11. Interface de visualisation de flux capturé par la caméra.

Chapitre 5 : Implémentation, résultats et comparaison

Cette interface dispose de cinq boutons dont les rôles sont décrits comme suit :

- Bouton « Charger Visages » : sert à afficher l'interface modélisant une simple visualisation de flux capturé sous forme d'une image couleur. Ce bouton permet de charger des visages.
- Bouton « Entraîner » : Ce bouton donne la main pour créer une base d'apprentissage.
- Bouton « Reconnaître » : affiche l'interface principale de notre système qui vous permet de tester la reconnaissance des expressions faciales par notre système.
- Bouton « Taux de performance » : Ce bouton sert à calculer les taux de reconnaissance des visages déjà capturés.
- Bouton « Ajouter une nouvelle personne » : Permet d'ajouter une nouvelle personne dans la base.
- Bouton « Vider la base » : Ce bouton sert à vider la base d'apprentissage.
- Bouton « Quitter » : Permet de quitter notre application.

Une fois le bouton « Entraîner » est sélectionné, la base est créée. Cette dernière action lance également la détection de visages dans l'image affichée, l'élimination des informations inutiles, le dessin des rectangles englobants le visage dans l'image couleur correspondante et le suivi de mouvement de ce visage dans la scène.

La construction d'une image d'apprentissage se déroule selon la chaîne suivante. D'abord, l'utilisateur clique sur le bouton « Charger Visages » avec l'expression de visage associée ensuite il clique sur le bouton « Entraîner » enfin il clique sur le bouton « Reconnaître ».

Le résultat de cette dernière action est la création de deux fichiers simultanément : un fichier « .bmp » présentant seulement le visage détecté et un fichier « .pcd » présentant le nuage de points correspondant à l'image de profondeur associée tout en n'extrayant que l'information utile relative dans notre cas seulement au nuage de points contenant le visage. Ces deux fichiers sont enregistrés dans un répertoire créé au préalable et dont la figure 5.12 présente un extrait.

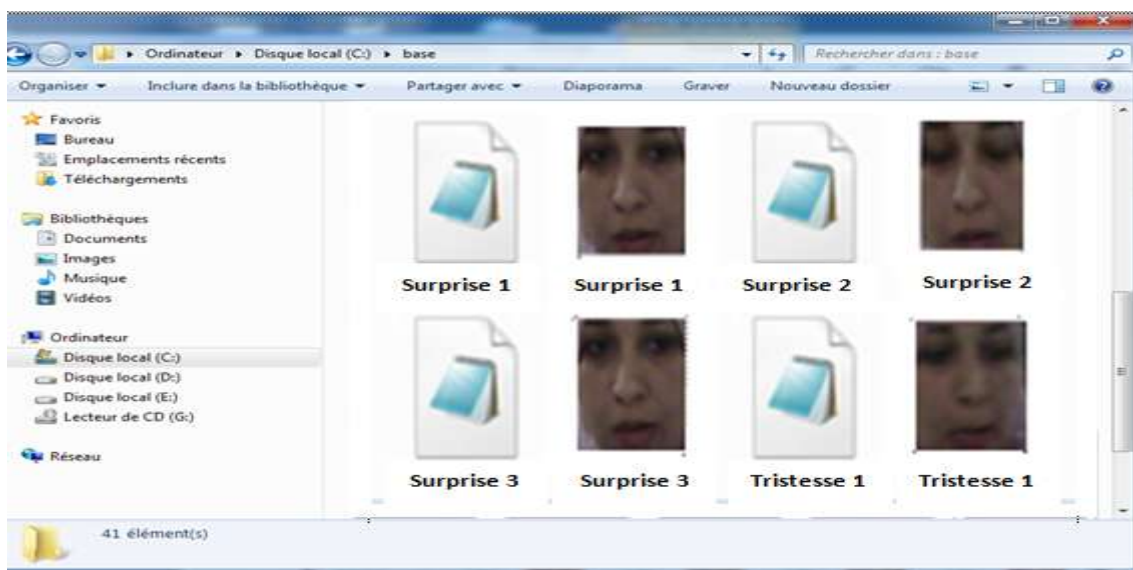


Figure 5.12. Extrait du répertoire « base ».

Chapitre 5 : Implémentation, résultats et comparaison

Pour répondre à l'exigence de la méthode « EigenFace », à savoir l'uniformité des images en taille, résolution et luminance, nous avons pris en considération cette contrainte en développant une première fonction servant à redimensionner la taille du visage à une taille bien déterminée que ce soit dans le cas où l'image d'entrée présente un visage plus grand ou plus petit par rapport à la taille fixée. Une deuxième fonction s'est chargée de convertir l'image couleur extraite en niveau de gris. Ces deux fonctions sont appelées lors de l'action générée par le bouton « Entraîner ».

S'agissant du bouton « Reconnaître », il sert à créer la liste des EigenExpressions relatives aux images de notre base, à calculer les poids de chaque image de notre base et enfin à sauvegarder la matrice résultante dans un fichier qui sera utilisé lors du processus de reconnaissance. Ce même bouton sert à actualiser également la matrice de poids et la liste des EigenExpressions une fois créée. Nous avons utilisé le package CVFaces pour l'implémentation de l'action relative à ce bouton tout en introduisant des modifications utiles pour notre cas d'application.

Une fois que nous ayons obtenu un fichier présentant les vecteurs de poids de chaque image de l'ensemble d'apprentissage, nous pouvons procéder au test de notre système via l'interface présentée dans la figure 5.13. Cette interface présente dans notre système, permet de visualiser le flux capturé par Kinect.

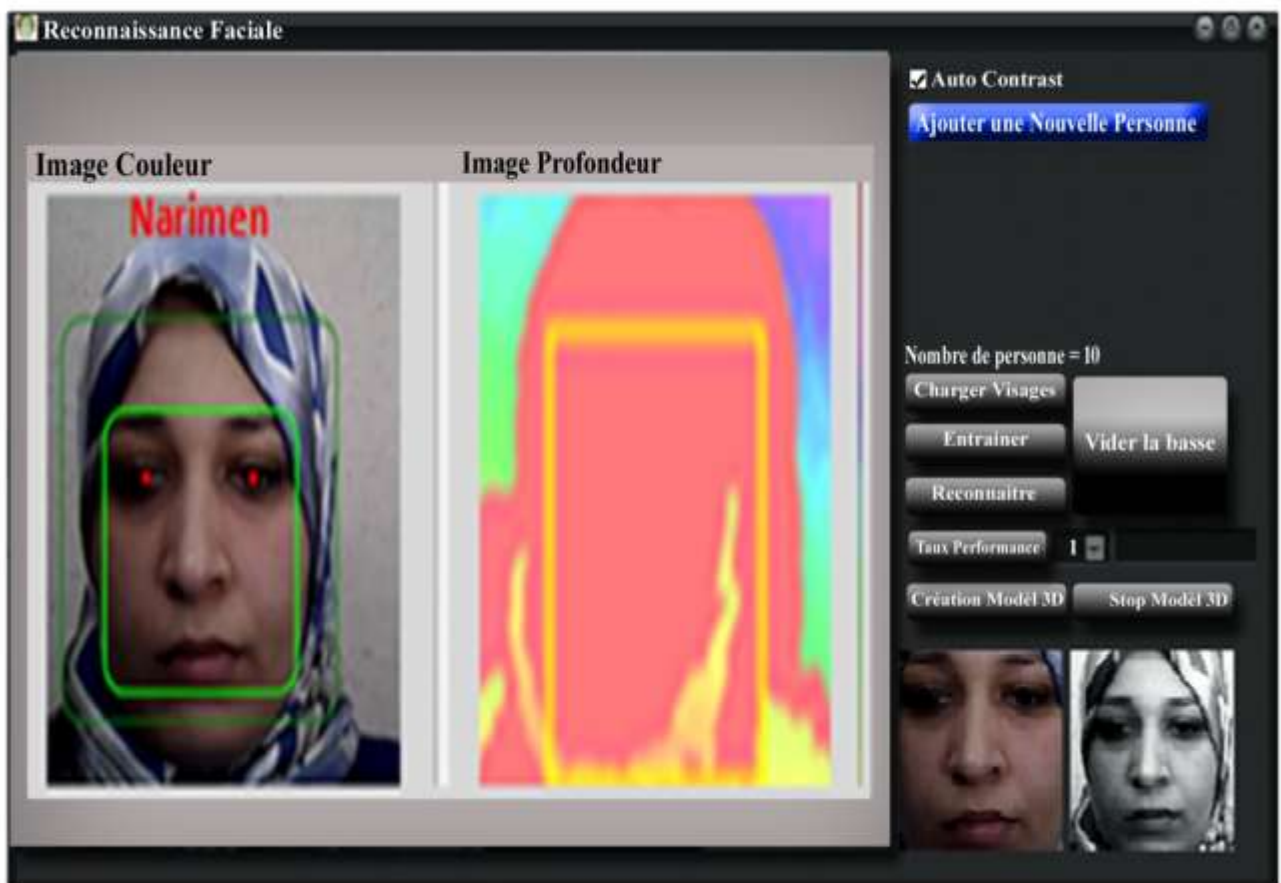


Figure 5.13. Flux de visualisation d'interface capturé par Kinect.

Chapitre 5 : Implémentation, résultats et comparaison

L'interface présente dans la figure 5.14 visualise, l'image couleur capturée, permettant le chargement du modèle 3D et l'expression associée.

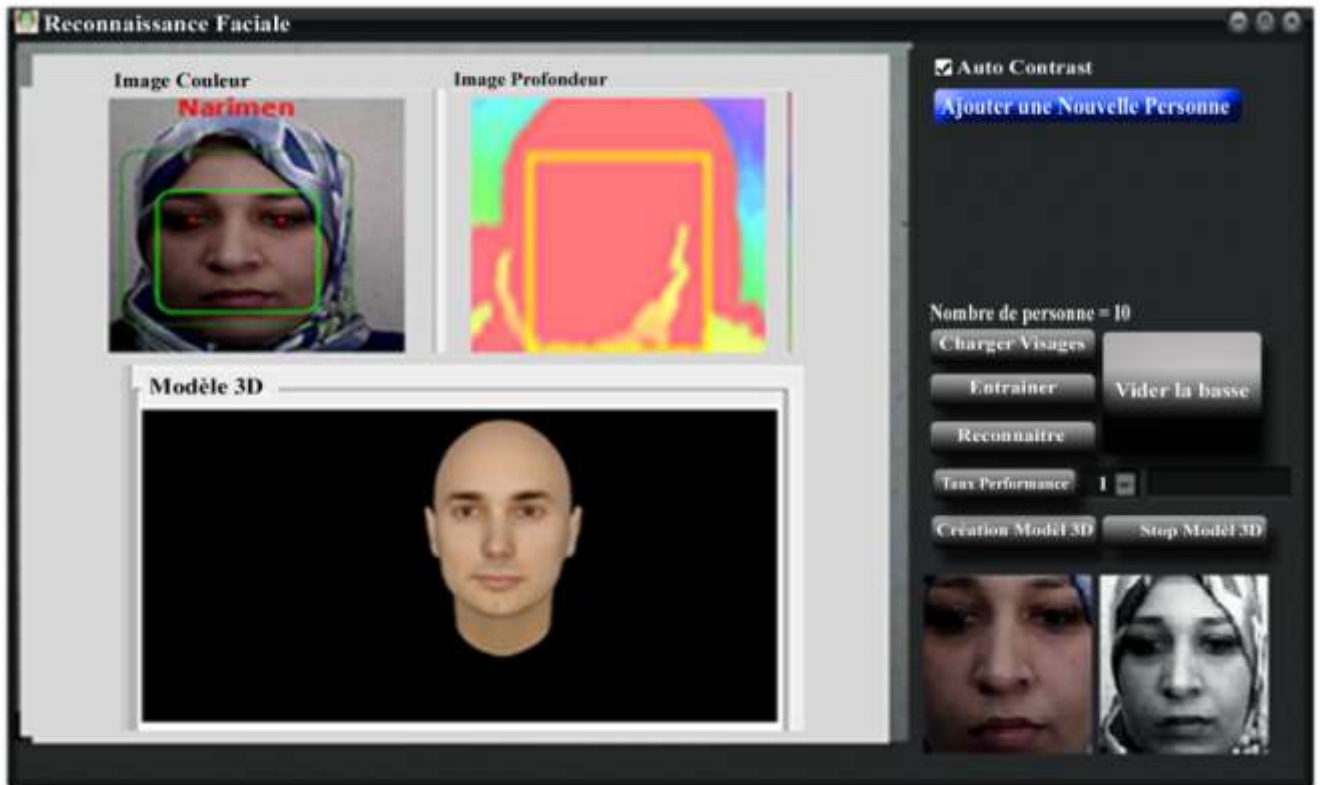


Figure 5.14. Interface permettant la reconnaissance du visage ainsi son expression détectée dans la scène.

Au cours du scénario d'exécution, cette interface peut s'afficher suite à un clic sur le bouton « Reconnaître ». En plus de l'affichage de l'interface, ce bouton sert à détecter le visage présent dans la scène, à redimensionner l'image pour respecter la taille fixée pour les images d'apprentissage et à la convertir en niveaux de gris tout en calculant le vecteur de poids correspondant. Il fait appel également à une fonction qui compare le vecteur de poids calculé avec la matrice sauvegardée dans le fichier en se servant de la distance euclidienne. Cette dernière fonction retourne le nom de la classe dont la colonne associée est la plus proche du vecteur de test. Nous nous servons aussi d'une autre fonction pour afficher l'étiquette de la classe identifiée sur l'interface et pour activer l'affichage du modèle 3D vers celle reconnue.

Le bouton « Création Modèle 3D » permet de créer des nuages de points 3D du visage capturé avec les dimensions générées par le capteur de la caméra tout en se limitant uniquement à notre zone d'intérêt, à savoir le visage. L'interface marque le visage ainsi que les différents traits faciaux par des boîtes englobantes, ce bouton ne rend pas un service exploité effectivement par notre application néanmoins, nous avons procédé à son implémentation dans le but de son exploitation dans la une étape ultérieure. Il est à noter également que le résultat de la segmentation est sauvegardé sur le disque dur en mentionnant la désignation du répertoire utilisé pour la sauvegarde.

5.2.3. Comparaison de notre système (AREFK) avec d'autres travaux

En 2014, Y. Sun, X. Wang et X. Tang [144] proposent d'apprendre un ensemble de représentations d'entités de haut niveau grâce à un apprentissage en profondeur, appelé DeepID IDentity (DeepID), pour la vérification du visage. Les fonctionnalités DeepID, lorsqu'elles sont apprises en tant que classificateurs permettent de reconnaître environ 10.000 identités de visages dans l'ensemble d'apprentissage. Les caractéristiques proposées sont extraites de diverses régions de la face pour former des représentations complémentaires et sur-complètes. Tous les classificateurs à la fine pointe de la technologie peuvent être appris à partir de ces représentations de haut niveau pour la vérification des visages. 97,45 % de précision de vérification sur LFW est obtenu avec seulement des faces faiblement alignées [144].

En 2016, D. Aneja et A. Colburn [145] ont proposé DeepExpr, une nouvelle approche de transfert d'expression de l'homme à de multiples personnages 3D stylisés. Ils forment d'abord deux réseaux neuronaux pour reconnaître l'expression des humains et des personnages stylisés indépendamment. Ensuite, ils ont utilisé des techniques d'apprentissage en profondeur comme un outil pour extraire des fonctionnalités utiles à partir de données brutes pour les visages humains et les personnages stylisés. Dans la dernière étape de la classification, l'algorithme tente de classer l'image du visage donnée en sept classes d'émotions de base différentes et ce, en utilisant des techniques d'apprentissage automatique. Les SVM sont les plus couramment utilisés pour les tâches de reconnaissance d'expressions faciales [145].

Notre système automatique reconnaissance d'expression de visage en utilisant Kinect (AREFK), on a les expressions automatiques sont reconnues avec le jeu de données Kinect (AREFK) d'images et de nuages de points contenant des informations relatives à 10 sujets différents avec chacune sept expressions différentes. (La figure 5.15 montre des exemples entre eux en utilisant la base de données FERF)

Plusieurs essais ont été provoqués pour chaque réaction, seulement quelques données ont été utilisées pour des raisons d'exactitude d'expression. Les faces sélectionnées ont été séparées en 70 expressions utilisées pour le processus d'apprentissage et 42 autres utilisées pour le test. L'acquisition des données a été recueillie dans une position fixe sous les mêmes conditions d'éclairage pour obtenir les meilleurs résultats de reconnaissance.



Figure 5.15. Facial Expression Research Group Base de données (FERF-DB) [202].

Le tableau 5.1 montre un exemple basé sur l'évaluation de 10 personnes avec 10 images/personne et sept expressions différentes retenues pour l'étape de test. Le meilleur pourcentage moyen trouvé est de 97 % correspondant à la 6e personne.

Tableau 5.1. Tableau de notre base de données de reconnaissance.

<i>Personnes</i>	<i>Base de données d'AREFK</i>						
	Naturel	Peur	Dégoût	Colère	Surprise	Joie	Tristesse
Personne 1	3	9	-	8	3	2	11
%	89%						
Personne 2	10	3	2	1	4	12	-
%	95 %						
Personne 3	-	2	7	1	-	8	1
%	93%						
Personne 4	1	5	-	6	1	3	-
%	87%						
Personne 5	-	-	1	6	9	1	9
%	82%						
Personne 6	13	-	-	-	2	7	-
%	97%						
Personne 7	2		3	5	1	-	4
%	82%						
Personne 8	-	5	2	3	10	9	-
%	79%						
Personne 9	2	1	-	3	6	-	7
%	85%						
Personne 10	3	-	-	4	7	-	2
%	86%						

Notre but avec l'exactitude de l'expression humaine est d'atteindre un bon résultat sur la reconnaissance d'expression qui soit assez proche aux résultats de l'état de l'art afin d'extraire des caractéristiques pertinentes relatives à une expression faciale.

Pour juger de l'efficacité de notre système, nous avons comparé AREFK à un expert humain. Nous avons demandé à l'expert et comparé leurs réponses avec les résultats de notre système AREFK en fonction de l'image de la requête humaine.

Notons qu'un nombre élevé de classifications incorrectes sont dues à des similitudes logiques dans la formulation de l'expression. En moyenne, l'état neutre, de joie et de surprise ont eu des performances de reconnaissance similaires. Par exemple, le dégoût est pratiquement similaire à la tristesse dans de nombreux cas. AREFK a montré une grande amélioration dans la reconnaissance de la tristesse et du dégoût. La peur a été confondue avec la surprise et l'état neutre en raison de la similitude sur les repères géométriques du visage montrant une bouche ouverte et les dispositions des yeux et le dégoût a été facilement confondu avec la colère. Bouche fermée, la tristesse et la colère ont été facilement confondues avec l'état de neutralité.

Le système AREFK est potentiellement acceptable par l'utilisateur ; il a obtenu un taux de reconnaissance des expressions de 79 % par rapport à un expert humain (90 %). A cause de ce constat, le choix automatique de la stratégie de reconnaissance ainsi que l'introduction de nos mesures de qualité dans le calcul du score doivent être optimisés. Les problèmes à résoudre sont essentiellement liés à la complexité du système actuel en termes de calcul et d'espace mémoire pour des intégrations futures.

Chapitre 5 : Implémentation, résultats et comparaison

Le tableau 5.2 résume la précision moyenne de reconnaissance des expressions pour chaque expression parmi toutes les expressions en utilisant un expert humain et notre système AREFK.

Tableau 5.2. Précision moyenne de reconnaissance d'expression (%) pour chaque expression à travers toute expression en utilisant un expert humain et AREFK.

Expression	Expert Humain	AREFK
Naturel	90	89
Peur	94	63
Dégout	83	76
Colère	96	77
Surprise	80	84
Joie	99	94
Tristesse	92	71

Le tableau 5.3 résume la performance de notre ensemble de données AREFK par rapport au jeu de données KiFaEx [203] et les résultats obtenus en combinant le jeu de données KifaEx & JAFFE [204] avec les deux autres jeux de données 64%, 81% respectivement.

Tableau 5.3. Montre la performance de notre base de données de reconnaissance et d'autres bases de données publiées.

Base de données	Précision moyenne de reconnaissance (%)
1- KiFaEx Dataset	64
2-KifaEx & JAFFE Dataset	81
3- AREFK Dataset "Notre base de donnée"	79

5.3 Contribution II : Implémentations, résultats et comparaison

Une partie du travail d'optimisation a néanmoins été effectuée afin de pouvoir obtenir des résultats dans des temps convenables pour les protocoles de la grande base de données FRGC.

L'efficacité des méthodes et l'approche proposée sont évaluées dans cette section. Tout d'abord, nous décrivons brièvement les ensembles de données sur lesquels les expériences ont été réalisées. Nous évaluons ensuite la vérification du visage 3D pour différentes méthodes d'extraction de caractéristiques binaires locales à différentes échelles. Enfin, la vérification du visage multimodale est effectuée et les résultats sont comparés avec ceux de la littérature scientifique.

5.3.1. Base de données FRGC V1.0 :

Pour cette expérimentation, nous avons utilisé le jeu de données FRGC v1.0. Les individus ont été acquis avec vue frontale au niveau de l'épaule, avec de très petites variations de pose.

Chapitre 5 : Implémentation, résultats et comparaison

Certaines analyses comprennent des occlusions du visage dues aux cheveux. Plus de détails sur l'ensemble de données FRGC peuvent être trouvés dans [150]. Les images ont été acquises avec le Scanner 3D « Minolta Vivid 910 » [23]. Les deux informations de couleur ainsi que les coordonnées 3D ne sont pas parfaitement capturées simultanément, car la bande de laser nécessite quelques secondes pour traverser le visage. La résolution de l'appareil photo Minolta est de taille 640×480, ce qui donne environ 300000 points d'échantillonnage possibles. Le nombre de points 3D sur une image frontale du visage prises par la caméra Minolta est d'environ 112000.

Des points supplémentaires proviennent de cheveux, vêtements et fond des objets Scans du visage sont donnés comme des matrices de taille 480x 640 de points 3D, avec un masque binaire indiquant les points valides de scan (par exemple, les points de premier plan correspondant typiquement à la tête et les épaules). En raison de différentes distances des sujets à partir du capteur au cours de l'acquisition, le nombre réel de points valables dans un balayage peut varier nous comparons la vérification unimodale de visage 3D dans la base de données FRGC v1.0, nous validons notre méthode proposée testons quatre méthodes d'extraction de caractéristiques binaires locales dans les données 3D étudiées dans ce chapitre.



Figure 5.16. Exemple de la base de données FRGC v1.0 [150].

5.3.2. Résultats

5.3.2.1. SIFT

A titre d'exemple, la figure 5.17 indique la répartition des valeurs de distance minimale entre les descripteurs de points clés SIFT détectés dans deux modèles considérés comme correspondant. La courbe verte (Intra) représente la répartition de la distance minimale à l'intérieur du sujet ainsi que la répartition des valeurs de la distance minimale entre les points clés détectés sur différents scans faciaux d'un même sujet. En revanche, la courbe rouge (Inter) représente la répartition de la distance minimale et inter représente la répartition des valeurs de la distance minimale entre les points clés détectés sur les balayages face à différents sujets. Comme prévu, la distribution intra-sujet obtient une distance minimale d'une forme (pic) à une valeur de portée beaucoup plus faible que la distribution de la distance minimale entre les sujets.

Cependant, les deux distributions ne sont pas assez séparées pour permettre une reconnaissance précise basée uniquement sur l'analyse statistique avec les valeurs de distance minimale entre les points clés détectés sur deux scans faciaux. Ceci est principalement incité par l'événement que les traits distinctifs du visage ne sont pas concentrés dans les régions restrictives de certains points du visage.

Les traits distinctifs du visage sont répartis sur tout le visage et, plutôt que d'être lié à la morphologie du visage à proximité d'un ensemble clairsemé de points clés du visage change de progrès depuis un point clé à un autre. Ceci reprend la définition des profils faciaux qui est discutée en détail dans la section suivante [198].

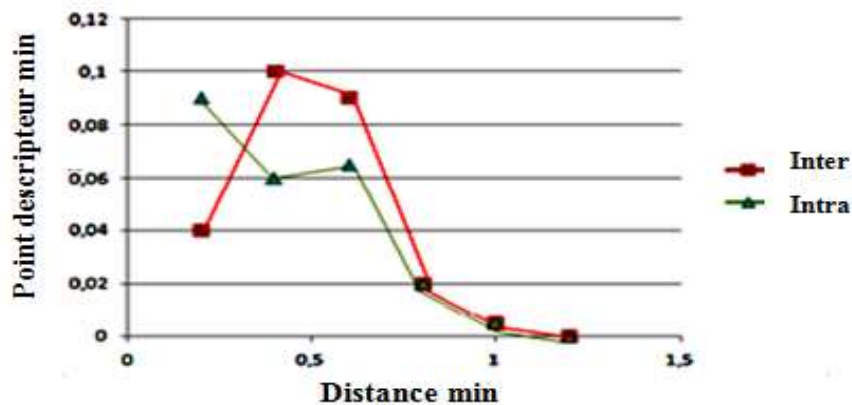


Figure. 5.17. Répartition des valeurs de la distance minimale entre les points clés des descripteurs SIFT détectés dans deux modèles des mêmes sujets (courbe verte) et pour différents sujets (courbe rouge). Les valeurs sont obtenues par la distance minimale moyenne sur toutes les paires de points clés dans le jeu de données FRGC v1.0.

Le cas, où la disponibilité d'une partie d'un visage uniquement affecte l'exactitude de la reconnaissance, est discuté dans ce qui suit. Dans la figure 5.18, on décrit les répartitions des valeurs de distance minimale entre les faces de profils qui sont générées à partir de différents points de clé, des modèles correspondants du même sujet (distance à l'intérieur du sujet) et divers sujets (distance inter-sujets). Ces courbes sont élaborées en calculant la moyenne des valeurs de la distance minimale de toutes les paires de balayages de la base de données FRGC v1.0 [198].

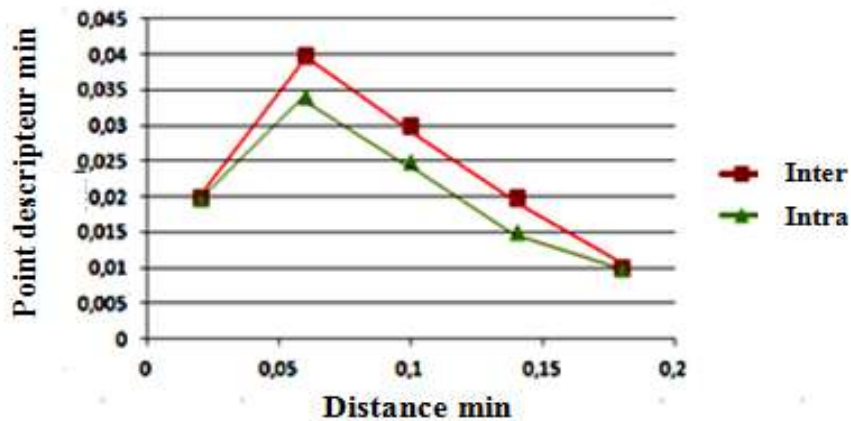


Figure. 5.18. Distribution des valeurs de distance minimale entre les points-clés détectés dans différents scans du même sujet (courbe verte) et différents sujets (courbe rouge). Ces courbes sont calculées en faisant la moyenne des valeurs de la distance minimale sur toutes les paires de balayages de la série de données FRGC v1.0.

Les images de profondeur sont obtenues avec les modèles de face 3D en fonction des changements de procédure et rapportées dans le prétraitement [198]. Selon la contribution proposée, les premiers points clés SIFT sont extraits d'une gamme d'images de toutes les faces. Cela se traduit par un nombre variable de points clés pour l'image, en fonction des caractéristiques spécifiques de la surface du visage. Pour chaque image de profondeur, seuls les 15 premiers points de clé (sélectionnés après avoir commandé tous les points de la plus grossière valeur de l'échelle) sont conservés et les profils faciaux sont calculés entre chaque paire de ces points clés.

La validité de la contribution proposée par rapport aux parties manquantes du visage a été testée en appliquant les paramètres expérimentaux suivants. Le motif (modèle) pour chaque sujet est sélectionné pour former la galerie (c'est-à-dire, nous avons choisi le modèle initial de chaque sujet comme modèle de galerie).

Ceci a été montré dans une galerie avec les modèles de référence 275, alors que tous les modèles ont été utilisés comme sondes et comparés à ceux de la galerie. Pour évaluer la précision de la solution proposée et différenciée par rapport à des parties de la face, chaque modèle de sondage est divisé en deux parties (parties gauche et droite avec observation du plan vertical parfait avec symétrie depuis la pointe) et les profils Dérivée d'une seule partie du visage différenciée à la galerie scans.

Exactement, deux expériences d'appariement séparées ont été réalisées, d'abord l'exercice depuis une sonde la partie gauche de chaque modèle de la face de la sonde, la seconde en utilisant la partie droite de chaque modèle de la face de la sonde. L'efficacité de la reconnaissance a été mesurée par le rang du taux de reconnaissance k et a présenté des courbes cumulatives de caractéristiques similaires (CMC) [198].

En particulier, la reconnaissance de l'expérience de rang k est réussie si la face de la galerie constituant la personne identique dans la présente sonde est classée en première position k de la liste de classement. Les courbes CMC sont évaluées à chaque valeur de k , le taux de k réussite des expériences.

La figure 5.19 présente les courbes CMC, par exemple où seuls les profils provenant d'une région du modèle du visage utilisé pour la reconnaissance (tracé pourpre correspondant au côté droit avec balayage du visage et tracé vert à gauche).

Pour estimer la robustesse de la contribution proposée, la figure 5.19 a également été notée avec un graphique de courbes bleues CMC pour le cas où tous les profils sont utilisés pour la reconnaissance. On peut observer qu'aucun mouvement pertinent n'émerge des profils employant le côté gauche ou droit du visage. Le taux de reconnaissance du rang 1 est voisin de 80% dans les deux cas. Il est tout aussi important de constater qu'il y a juste une légère diminution de performance par rapport au cas où les profils faciaux de la face globale sont utilisés dans la correspondance [198].

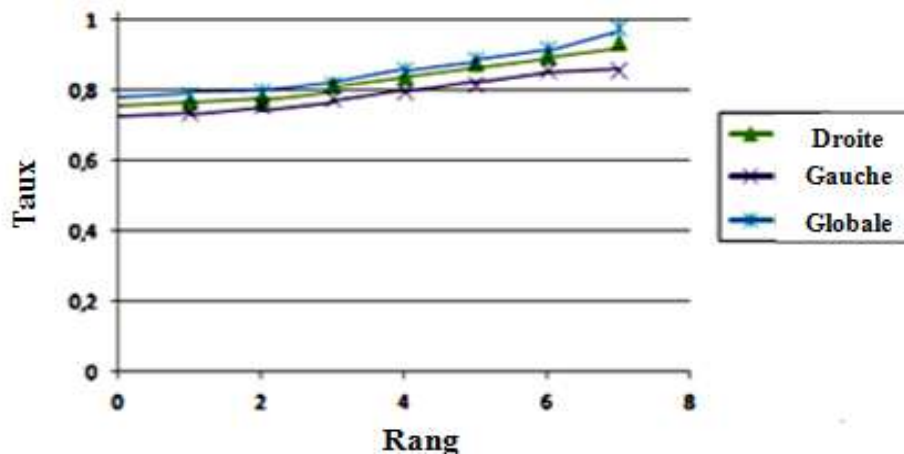


Figure. 5.19. Tracés des courbes CMC obtenues en utilisant des modèles de sonde à partir de profils de points clés sur le visage (bleu) côté gauche du visage (pourpre) et sur le côté droit du visage (vert).

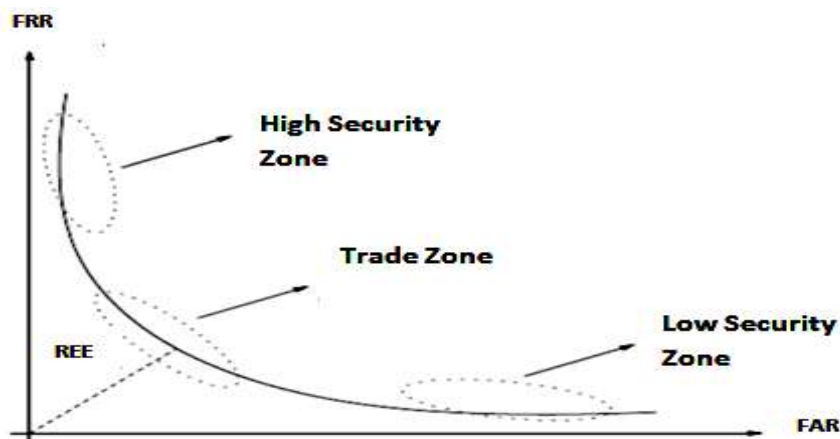


Figure. 5.20. Courbe ROC [205].

Cependant, les courbes ROC (Receiver Operating Characteristic) mesurent la performance d'un système d'authentification (utilise également le terme reconnaissance 1 : 1). La courbe ROC (Figure 5.20) fournit une courbe FRC (False Rejection Rate - FRR) contre FSA (Faux Taux d'Acceptation) (FAR) est une autre façon de montrer les données et peut être calculée à partir des FAR et FRR. Une étude comparative spécifiant la relation entre les courbes CMC et ROC peut être trouvée dans [105].

5.3.2.2 Méthode d'extraction de caractéristiques LBP

Dans cette partie, nous utilisons la méthode locale LBP.

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

- 1) *Taux de Faux Rejet (TFR) ou (False Reject Rate, FRR)* : ce taux représente le pourcentage de personnes censées être reconnues, mais qui sont rejetées par le système.
- 2) *Taux de Fausse Acceptation (TFA) ou (False Accept Rate, FAR)* : ce taux représente le pourcentage de personnes censées ne pas être reconnues, mais qui sont tout de même acceptées par le système.
- 3) *Taux d'Égale Erreur TEE ou (Equal Error Rate, EER)* : Ce taux est calculé à partir des deux premiers critères et constitue un point courant de mesure de performance. Ce point correspond à l'endroit où $TFR = TFA$, c'est-à-dire le meilleur compromis entre les faux rejets et les fausses acceptations.
- 4) *Taux de reconnaissance (TR) ou (Recognition Rate, RR)* : Cela mesure le pourcentage d'entrées correctement identifiées.
- 5) *Taux de vérification (TV)* : $TV = 100 - TFR - TFA$.

Le tableau 5.4 montre le taux d'erreur dans l'ensemble d'évaluation et de test par cette méthode d'extraction de caractéristiques. Le nombre de points voisins (P) varie de 8 à 24 points et la valeur du rayon (R) varie d'un à quatre pixels.

La méthode LBP pour l'information 3D donne plus de performances pour les quatre valeurs du rayon (R), (Voir chapitre 4, section 4.4.2.3 « Image de profondeur et d'intensité »). La fusion des quatre valeurs du rayon (LBP) améliore les performances des informations 3D de visage de la base de données FRGC v1.0, avec un TEE = 2.27 % dans l'ensemble d'évaluation et TV = 93,25 % dans l'ensemble de tests. Aussi pour l'information 2D de visage, avec TEE = 1.89 % dans l'ensemble d'évaluation et TV = 93,23% dans l'ensemble de tests.

Tableau 5.4. Performance de la méthode d'extraction de caractéristiques LBP dans l'ensemble d'évaluation et de test.

(P,R)	Image d'intensité (2D)		Image de profondeur (3D)	
	Eval TEE	Test TV	Eval TEE	Test TV
(8,1)	6.90	82.07	7.13	89.04
(16,2)	3.45	88.14	5.80	90.19
(24,3)	5.18	89.96	6.85	89.17
(24,4)	4.1	90.29	5.94	91.21
<i>1+2+3+4</i>	1.89	93.23	2.27	93.25

5.3.2.3 LBP +SIFT

Le tableau 5.5 indique le taux d'erreur dans l'ensemble d'évaluation et de test pour l'extraction de caractéristiques par LBP+SIFT, en utilisant la base de données FRGC v1.0. Le nombre de points voisins (P) varie de 8 à 24 points et la valeur du rayon (R) varie entre un et quatre pixels. L'adoption de cette approche (LBP+SIFT) nous semble un bon compromis entre performance et complexité algorithmique.

Le tableau montre que la fusion de quatre LBP (R=1, 2, 3, 4 et P=8, 16, 24) plus SIFT donne le meilleur résultat avec un TEE =2.67 % et TV= 94. 98, pour l'information 3D de visage de la base de données FRGC v1.0.

Tableau 5.5. Performance de la méthode d'extraction de caractéristiques LBP + SIFT dans l'ensemble d'évaluation et de test.

(P,R)+SIFT	Image d'intensité (2D)		Image de profondeur (3D)	
	Eval TEE	Test TV	Eval TEE	Test TV
<i>(8,1)+SIFT</i>	8.97	81.67	6.73	88.44
<i>(16,2)+SIFT</i>	6.09	87.47	6.30	89.89
<i>(24,3)+SIFT</i>	4.18	91.66	7.25	88.7
<i>(24,4)+SIFT</i>	3.7	89.49	6.34	90.51
<i>(1+2+3+4)+SIFT</i>	2.91	92.94	2.67	94.98

5.3.3. Comparaison des algorithmes de réduction de l'espace de données

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

Nous avons étudié 3 algorithmes de réduction d'espace de données pour choisir le meilleur algorithme. Nous utilisons plusieurs algorithmes de réduction d'espace. On peut légitimement se poser la question : quel algorithme choisir ? Pour répondre à cette question, nous avons effectué nos expériences sur tous ces algorithmes pour le choix du meilleur algorithme.

La figure 5.21 nous montre que la méthode SIFT+LBP est plus performante que la SIFT suivie de la méthode LBP et l'ACP en employant un nombre restreint de caractéristiques (vecteurs propres) dans notre système de vérification du visage 3D en utilisant la base de données FRGC v1.0.

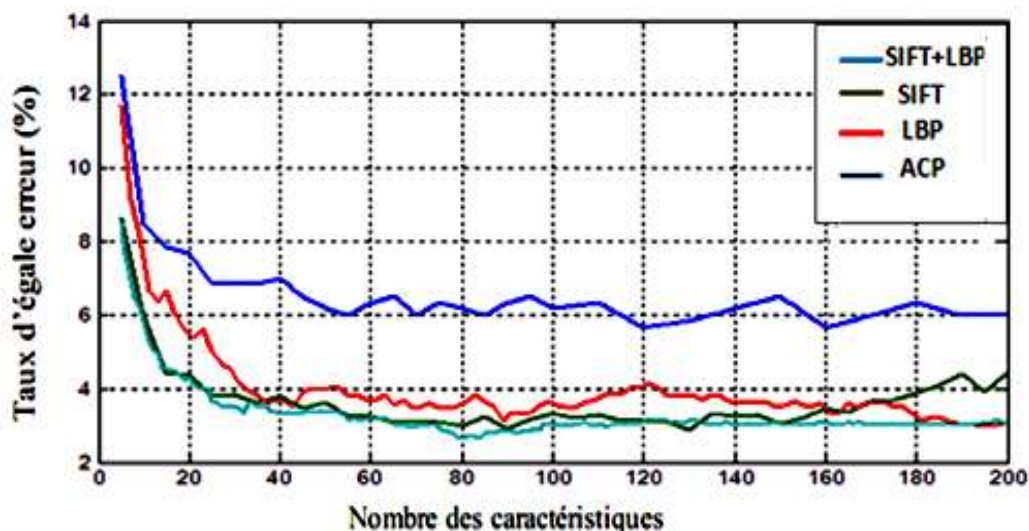


Figure 5.21. Comparaison des TEE pour SIFT+LBP, SIFT, LBP, ACP dans l'ensemble d'évaluation.

Le tableau 5.6 nous montre bien que le meilleur résultat est obtenu avec SIFT+LBP avec un TEE=2.67% et un TV=94.98% en utilisant seulement 80 paramètres caractéristiques.

Nous remarquons aussi que les TV et TEE obtenus par la ACP et l'LBP sont assez proches. Alors que le mauvais résultat est obtenu par ACP appliquée seule. C'est un résultat bien évident puisque cette méthode est gourmande en termes de volume de données et n'assure pas une discrimination optimale. L'algorithme de réduction de l'espace de données retenu reste le SIFT+LBP.

Tableau 5.6. Performances des méthodes SIFT+LBP, SIFT, LBP, ACP.

Méthodes	Ensemble Évaluation		Ensemble de tests		Nombre de caractéristiques
	TEE %	TFR%	TFA%	TV%	
ACP	5.7	2.7	8.1	89.2	200
SIFT	2.83	2	4.12	79.98	90
SIFT+LBP	2.67	2	3.66	94.98	80
LBP	2.27	2.1	4.3	93.27	195

Malgré des résultats satisfaisants pour les systèmes d'authentification de visages obtenus dans cette partie, les taux d'erreur associés à des systèmes biométriques unimodaux restent relativement élevés. La complexité algorithmique et les longs délais d'exécution de ces algorithmes (développés à l'INT et non optimisés), ne nous permettent pas d'exploiter de façon efficace ces approches sur des grandes bases de données comme FRGC.

5.3.4. Comparaison avec d'autres travaux

Le tableau 5.7 résume les performances de la méthode proposée et d'autres résultats publiés de l'indice de référence, de sorte que nous pouvons noter que nos résultats sont meilleurs que les autres méthodes, donnant des résultats satisfaisants.

Tableau 5.7. Montre la performance de la méthode proposée et d'autres résultats publiés de l'indice de référence.

Approche	Caractéristiques
1-Faltemier and al. [206]	- Sensible, sauf les parties manquantes (fusion difficile des résultats).
2-Aly UZ, G Okberk [207]	- Traitement d'occlusion seulement.
3-Berretti, Del Bimbo [198]	- Taux= 70% pour un rang k = 1 sans données de classification de chaque partie du visage.
4- Notre approche	- Taux= 80% pour un rang k = 1 avec des données de classification de chaque partie du visage + Traitements d'occlusion et parties manquantes ; - L'utilisation de quatre LBP + SIFT améliore nos résultats avec VR = 94,98%.

Les performances individuelles des enregistrements de travail de reconnaissance faciale 3D avec des parties manquantes sont celles de Faltemier et al. [206]. Le travail d'Aly UZ et G Okberk [207] n'a traité que des occlusions faciales qui ont mentionné un taux de reconnaissance de rang 1

Chapitre 5 : Implémentation, résultats et comparaison

en utilisant l'ensemble de données FRGC de 88% et 89,2%, séparément pour les régions du côté gauche et du côté droit du visage. Cependant, le réglage expérimental appliqué dans ces évaluations suppose que les régions ne sont aucunement affectées par une partie manquante correspondante et qu'il est difficile de fusionner les résultats.

S. Berretti, A. Del Bimbo et P. Pala [198] ont proposé une approche pour la reconnaissance faciale 3D basée sur les propriétés locales du visage. L'approche est motivée par le fait que l'information capturée par le descripteur SIFT en relation avec les points clés dispersés du visage est également locale pour capturer les caractéristiques de la face 3D qui prennent en charge la reconnaissance de visage spécifique. Dans la différenciation de deux faces, les descripteurs SIFT sont appariés pour calculer la similarité entre les paires de points clés identifiés sur les deux images de distance. Ensuite, la distance entre les deux faces est dérivée en constituant les distances individuelles entre les profils faciaux qui proviennent des points clés égaux. La méthode proposée est testée à l'aide de l'ensemble de données FRGC v1.0.

Pour estimer la précision de la reconnaissance en présence de parties manquantes, ils ont adopté un protocole d'évaluation selon lequel seuls les profils issus d'une partie du balayage de la sonde sont comparés aux profils des balayages de la galerie. Le taux de reconnaissance de rang 1 est proche de 70 %.

Après le travail de [208, 198, 206], nous avons proposé une approche de reconnaissance faciale tridimensionnelle basée sur les propriétés locales du visage. Nous avons également adopté le principe d'évaluation selon lequel seuls les profils issus d'une partie de la sonde scan sont comparés aux profils de balayage de la galerie avec les données de classification de chaque partie du visage en utilisant la méthode SVM. Le taux de reconnaissance de rang 1 est proche de 80 %. Ensuite, nous avons amélioré nos résultats en utilisant pour l'extraction de caractéristiques quatre LBP + SIFT avec VR = 94,98 %.

Le tableau 5.8 résume la performance de la méthode proposée et la comparaison avec d'autres méthodes que nous avons mises en œuvre, puis nous avons testé et constaté que les résultats de l'opérateur SIFT sur les images LBP_{P, R} sont supérieurs à d'autres méthodes, donnant des résultats satisfaisants.

Tableau 5.8. Classification des Résultats pour le jeu de données FRGC v1.0 avec la méthode SIFT + LBP et comparaison avec nos autres méthodes utilisées.

Méthodes d'extraction de caractéristiques	Taux de reconnaissance (%)
1- PCA et Hu + SVM	89.2
2- SIFT + SVM	79.98
3-LBP + SVM	93.27
4- SIFT et LBP + SVM	94.98

5.4. Conclusion

Dans ce chapitre, nous avons présenté nos résultats en deux temps. Dans une première partie, nous avons utilisé un système capable d'identifier une expression faciale en temps réel avec un coût de production raisonnable. Nous proposons de tirer profit de la performance atteinte par les techniques 2D et de corriger les lacunes par l'exploitation de la géométrie tridimensionnelle du visage grâce à la carte de profondeur offerte par le capteur Kinect.

Dans la seconde partie, selon l'approche que nous avons proposée, nous avons également suivi le principe d'évaluation lorsque seuls les profils issus d'une partie du visage scanné sont utilisés. Ceci est réalisé dans le but de traiter des problèmes majeurs dans le domaine de la reconnaissance faciale tridimensionnelle et multimodale en tenant compte des variations de l'éclairage, de l'expression, de petits mouvements tels la rotation de la tête et les changements dans l'expression du visage combinés avec les changements d'éclairage. L'application utilise les bases de données FRGC v1.0 des visages 3D. Le taux de reconnaissance de rang 1 est proche de 80 %.

Nous nous sommes ensuite attelés à l'amélioration de nos résultats avec un système multimodal d'authentification du visage qui offre de bonnes performances. Tout d'abord, nous avons corrigé la rotation de la tête par l'algorithme ICP (Iterative Closest Point), puis présenté notre algorithme de prétraitement. Pour l'extraction des caractéristiques, quatre méthodes locales ont été utilisées, à savoir : ACP, SIFT, LBP, SIFT+LBP. La fusion de quatre LBP ($R = 1, 2, 3, 4$ et $P = 8, 16, 24$) plus SIFT donne le meilleur résultat avec $TEE = 2,67 \%$ et $TV = 94,98\%$.

Conclusion générale et perspectives

Ce travail est divisé en deux parties ; Tout au long de la première contribution nous avons organisé notre effort sur trois étapes :

D'abord, nous nous sommes intéressés à la reconnaissance des expressions faciales en utilisant la méthode « MB_LBP ». Dans ce cadre, nous avons consacré une première sous-étape à la détection du visage afin de différencier sa présence/absence, de poursuivre sa localisation par un rectangle englobant et de reconnaître les différents traits d'expressions faciales par des rectangles.

Dans la deuxième étape, nous avons utilisé la matrice des poids fournie par l'étape d'apprentissage pour déterminer automatiquement l'expression faciale exprimée par l'utilisateur.

Pour cela, nous avons présenté les détails concernant l'implémentation de notre solution. Dans la première section, nous avons présenté les outils matériels et logiciels avec lesquelles nous avons travaillé. Nous avons passé à la présentation des différentes fonctionnalités de notre système avant d'exposer et commenter, dans la dernière section, les différentes interfaces composant l'application produit de notre travail, développée en C++. Bien que la solution proposée puisse améliorer et faciliter la recherche dans le domaine de l'interaction Homme-Machine puis qu'il s'agit d'une approche totalement automatique permettant l'extraction de visages et leurs caractéristiques, la reconnaissance des expressions faciales, quelques approfondissements sont utiles.

Comme première perspective d'évolution du travail réalisé, nous envisageons de généraliser notre modèle pour couvrir une panoplie plus large d'expressions faciales. L'étude d'autres techniques de reconnaissance des expressions faciales notamment à travers le flux d'images de profondeur et les modèles 3D qui peuvent en résulter, ainsi que la fusion des données en provenance de différentes sources 2D et 3D. Nous envisageons également de faire une étude dimensionnelle des expressions faciales.

Pour des développements futurs, notre intérêt s'est porté vers la reconstruction du modèle 3D du visage. Nous voulons récupérer le nuage de points de la scène réelle à partir de la carte de profondeur. Ainsi, nous nous sommes concentrés sur la fusion des résultats des deux étapes de reconnaissance et de reconstruction pour atteindre un objectif double à savoir l'implémentation d'un système convenable à la reconnaissance des expressions et le mappage du résultat de la reconnaissance vers un modèle 3D virtuel. En nous basant sur les travaux de Zollhöfer et al [42] et de Weise et al [43], desquelles nous nous sommes inspirés pour réaliser la reconstruction et l'animation de modèle 3D.

Tout au long de la deuxième contribution nous avons :

Réalisé une étude expérimentale de l'authentification de visages sur la base de données FRGC v1.0 en commençant par la comparaison des algorithmes de réduction de l'espace présentée dans les chapitres précédents. Pour la comparaison de méthodes de réduction de l'espace de données, nous

Conclusion générale et perspectives

avons utilisé quatre méthodes (Analyse en Composantes Principales (ACP), Scale Invariant Feature Transform (SIFT), la méthode Local Binary Pattern LBP) et la fusion des deux méthodes (SIFT+LBP)). Dans cette étude, nous déduisons que la méthode SIFT+LBP est celle qui offre les résultats les plus probants. Enfin, nous avons sélectionné les meilleurs systèmes d'authentification de visages pour construire un système multi-algorithmes dans la deuxième partie de cette section.

Malgré des résultats très intéressants, nous nous sommes rendus compte que nous atteignons des limites en ne considérant que la base de données contrôlée (FRGC v1.0) qui contient des variations d'expressions, modifications d'éclairage, rotation de la tête, bruits introduits par le capteur, occlusions ..., carences que nous devons pallier si l'on souhaite améliorer les performances.

Nous avons étudié les limitations des systèmes biométriques lorsqu'ils n'utilisent qu'une seule modalité biométrique. Ensuite, nous présentons une façon de réduire les limitations des systèmes biométriques monomodaux en combinant plusieurs systèmes menant ainsi à la biométrie multimodale.

Les systèmes multimodaux peuvent être de différentes natures. Pour cela, nous avons étudié les différents types de combinaisons de modalités possibles, mais aussi les niveaux de fusion qui peuvent être utilisés dans un système multimodal. Nous présentons donc le type de multimodalité et le niveau de fusion.

Parmi les divers niveaux de fusion existants, nous avons choisi de travailler au niveau score car il offre le meilleur compromis entre la richesse de l'information et la facilité d'implémentation. Un état de l'art de la fusion de scores, la théorie des méthodes de fusion ainsi que six méthodes de normalisation de scores utilisées ont été présentées.

Enfin, nous avons étudié, dans les résultats expérimentaux, un système multi-algorithmes d'authentification de visages. Ce qui rend nécessaire l'étape de normalisation des scores.

Dans la section de comparaison des méthodes de classification de scores, nous proposons d'utiliser les méthodes de normalisation de scores avant la classification ; c'est ce qui fait l'originalité de notre travail. D'après toute cette étude et les résultats obtenus, nous pouvons affirmer que les méthodes de normalisation de scores améliorent la performance en général pour la méthode de classification de scores utilisé. La méthode de classification de scores non linéaires SVM et donne des résultats équivalents. Cependant, le SVM présente une simplicité.

Les travaux futurs porteront sur l'utilisation d'un modèle pour estimer l'information mutuelle entre les problèmes et les points clés/profils afin d'identifier les points clés/profils qui peuvent être associés de manière fiable à des modèles d'un même sujet et nous cherchons à utiliser la base de données FRGC V2.0. La multimodalité sera abordée à partir de l'apparence de l'angle, plus comme une forme de première manière ; et divers scores sur diverses parties du visage (par exemple yeux, bouche) se combinant d'une autre manière.

REFERENCES

- [1] S. Hamida. " L'extraction de caractéristiques pour l'analyse biométrique tridimensionnelle d'un visage ". Thèse de Mastère 2, Département d'informatique, Université de Mohamed Khider Biskra, Juin 2010.
- [2] M. Chassé. " La biométrie au Québec : Les enjeux ". Commission d'accès à l'information au Québec, Juillet 2002.
- [3] C.L. Tisse. " Contribution à la vérification biométrique de personnes par reconnaissance de l'iris ". Thèse de doctorat, Université de Montpellier II, 28 octobre 2003.
- [4] A. Ouamane. " Reconnaissance Biométrique par Fusion Multimodale du Visage 2D et 3D". Thèse de doctorat en sciences en Electronique, Université de Mohamed Khider Biskra, 11 Juin 2015.
- [5] A. Jain, L. Hong, S. Pankanti et R. Bolle. "An Identity Authentication System Using Fingerprints ". Proceedings of the IEEE ISSN 0018-9219, volume 85, Numéro 9. 1997.
- [6] P. Phillips, H. Wechsler, J. Huang et P. Rauss. "The FERET Database and Evaluation Procedure for Face Recognition Algorithms". Image and Vision Computing, Volume 16, pages 295-306, 1998.
- [7] F. Perronnin et J. Dugelay. " Introduction à la Biométrie Authentification des Individus par Traitement Audio-Vidéo ". Revue Traitement du Signal, Volume 19, numéro 4, 2002.
- [8] S. G. Ababsa. "Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D ". Thèse de doctorat, Université Evry Val d'Essonne. 03 octobre 2008.
- [9] C. Beumier. "Authentification d'identité par analyse tridimensionnelle du visage ". Thèse de doctorat, Université de Paris, Décembre 2003.
- [10] N. George. "MEG/EEG et reconnaissance des visages ", CNRS UPR 640 LENA, Hôpital de la Salpêtrière, Paris.
- [11] P. J. Phillips, P. Grother, R.J. Micheals, D.M. Black-Burn, E Tabassi et J.M. Bone. "Face recognition vendor test 2002". Evaluation Report. Technical Report 6965, NIST, Mars 2003.
- [12] M. Marzouki. " Enjeux des techniques de biométrie ". Troisième conférence internationale des commissaires à la protection des données, vie privée, droits de l'homme, septembre 2001. Paris, France.

Références

- [13] M. T. Laskri et D. Chefrour. " Who_Is : système d'identification des visages humains ". Revue Internationale ARIMA (INRIA), Volume 1, pages 39 à 61, France 2003. ISSN 1638-5713.
- [14] K. I. Chang, K. W Bowyer et P. J. Flynn. " An Evaluation of Multi-Modal 2D+3D Face Biometrics", IEEE Transactions on pattern Analysis and Machine Intelligence, 27 (4): 619-624, Avril 2005.
- [15] S. Romdhani, V. Blanz, C. Basso et T. Vetter. " Morphable Models of Faces" Dans S.Z. LI et A.K. JAIN, éditeurs, Handbook of Face recognition, Chapitre 10, Springer - Verlag Reidel, Dordrecht, 2004. 33 pages.
- [16] M. Visani, C. Garcia, J-M. Jolion. " Une nouvelle méthode de représentation des visages pour leur reconnaissance : l'analyse discriminante bilinéaire ", France Telecom Division R&D TECH / IRIS. Laboratoire LIRIS, INSA Lyon.
- [17] D. Blackburn, M. Bone, P. J Phillips. "Face recognition vendor test 2000". Tech. Rep. <http://www.frvt.org>, 2001.
- [18] P. Sinha, B. J. Balas, Y. Ostrovsky and R. Russell. "Face recognition by humans". Face Processing: Advanced Modeling and Methods. Academic Press, San Diego, 2006.
- [19] Minoltavivid910url.<http://www.konicaminolta.com/instruments/products/3d/noncontact/vivid910/features.html>, 2011.
- [20] Inspeck mega capturoor ii digitizer url. <http://www.inspeck.com>, 2011.
- [21] Kinect url. <http://www.xbox.com/en-GB/kinect>, 2011.
- [22] S. Zafeiriou, M. Hansen, G. Atkinson, V. Argyriou, M. Petrou, M. Smith and L. Smith. "The photoface database". In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Workshop on Biometrics) (CVPRW2011), 2011.
- [23] P. J. Phillips, P. J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. "Overview of the face recognition grand challenge". in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1. Washington, DC: IEEE Computer Society, 2005, pp. 947-954.
- [24] L. Yin, X. Wei, Y. Sun, J. Wang and M. J. Rosato. "A 3D facial expression database for facial behavior research". In International Conference on Automatic Face and Gesture Recognition, pp. 211-216, 2006
- [25] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur and L. Akarun. "Bosphorus database for 3D face analysis". In COST Workshop on Biometrics and Identity Management, 2008
- [26] C. Xu, T. Tan, S. Z. Li, Y. Wang, and C. Zhong. "Learning effective intrinsic features to boost 3D-based face recognition". in Proc. 9th Eur. Conf. Comput. Vis. (ECCV), Part II, LNCS 3952, pp. 416-427, May 2006.

Références

- [27] T. Heseltine. "Face recognition: two-dimensional and three-dimensional techniques". PhD thesis, The University of York, 2005.
- [28] V. Blanz and T. Vetter. "A morphable model for the synthesis of 3D faces". In Annual Conference on Computer Graphics, pp. 187-194, 1999.
- [29] L. D. Ruderman. "The statistics of natural images". *Network: Computation in Neural Systems*, pp. 517-548, 1994.
- [30] P.S. Penev and J.J. Atick. "Local feature analysis: A general statistical theory for object representation", *Network : Computation in Neural Systems*, vol. 7, no. 3, pp. 477-500, 1996.
- [31] D. Saigaa. "Contribution à l'authentification d'individus par reconnaissance de visages", thèse de Doctorat d'état en automatique, Université Mohamed Khider, Biskra, Algérie, Novembre 2006.
- [32] A. Mehrabian. "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament", *Current Psychology*, volume 14 (4), pages 261-292, 1996.
- [33] X. Tan, S. Chen, Z. H. Zhou et F. Zhang. "Face recognition from a single image per person: A survey". *Pattern Recognition*, Volume 39, numéro 9, Pages 1725-1745, Septembre 2006.
- [34] N. Morizet. "Reconnaissance Biométrique par Fusion Multimodale du Visage et de l'Iris". Thèse de l'Ecole Nationale Supérieure des Télécommunications Spécialité : Signal et Images, 2009.
- [35] M. Turk et A. Pentland. "Eigenfaces for Recognition ". *Journal of Cognitive Neuro science*, Volume 3, Numéro.1, pages71-86, 1991.
- [36] R. Szeliski. "Computer Vision: Algorithms and Applications", livre (website <http://szeliski.org/Book>), Springer, Août 2010.
- [37] B. Ans, J. Héroult, and C. Jutten. "Adaptive neural architectures: detection of primitives". In *COGNITIVA*, pp. 593-597, 1985.
- [38] J. Héroult, B. Ans, and C. Jutten. "Circuits neuronaux à synapses modifiables : Décodage de messages composites par apprentissage non supervisé". In *Comptes Rendus de l'Académie des Sciences*, pp. 525–528, 1984.
- [39] P. Comon. "Independent component analysis - a new concept", *Signal Processing*, vol. 36, no. 3, pp. 287-314, 1994.
- [40] C. Jutten and J. Héroult. "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture". *Signal Processing*, vol. 24, pp. 1-10, 1991.
- [41] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. "Face recognition by independent component analysis". *Transactions on Neural Networks*, August 15 2002.

Références

- [42] W. S. Torgerson. "Multidimensional scaling". *Psychometrica*, Vol. 17, pp. 401-419, 1952.
- [43] D. D. Lee and H. S. Seung. "Learning the parts of objects by non-negative matrix factorization". *Nature*, vol. 401, pp. 788-791, October 1999.
- [44] B. Moghaddam et A. Pentland. "Probabilistic visual learning for object representation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 19, Numéro 7, pages 696-710, July 1997.
- [45] P. Belhumeur, J. Hespanha et D. Kriegman. "Eigenfaces vs. fisherfaces: recognition using class specific linear projection". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 19, No. 7. July 1997.
- [46] J. Lu, K.N. Plataniotis et A.N. Venetsanopoulos. "Face recognition using kernel direct discriminant analysis algorithms". *IEEE Transactions on Neural Networks*. Volume 14, No. 1, January 2003.
- [47] R. A Fisher. "The use of multiple measures in taxonomic problems". *Ann. Eugenics*, 7:179-188, 1936.
- [48] Daniel L. Swets and J. Weng. "Using discriminant eigenfeatures for image retrieval". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, No. 8, pp. 831-836, 1996.
- [49] B. Scholkopf, A. Smola, and K. Muller. "Nonlinear component analysis as a kernel eigenvalue problem". *Neural Computation*, pp. 1299-1319, 1998.
- [50] J. Shawe-Taylor and N. Cristianini. "Kernel Methods for Pattern Analysis". CUP, June 2004.
- [51] L. Wiskott, J. M. Fellous, N. Kuiger, and C. von der Malsburg. "Face recognition by elastic bunch graph matching". *IEEE Trans. PAMI*, vol. 19, no. 7, pp. 775-779, July 1997.
- [52] Ngoc Son Vu. "Contributions à la reconnaissance de visages à partir d'une seule image et dans un contexte non-contrôlé ", Thèse de L'Université de Grenoble, 2010.
- [53] B. S. Manjunath, R. Chellappa, and C. von der Malsburg. "A feature based approach to face recognition". In *CVPR*, 1992.
- [54] R. Brunelli and T. Poggio. "Face recognition : features versus templates". *IEEE Trans. PAMI*, vol. 15, no. 10, pp. 1042-1052, 1993.
- [55] P. Viola and M.J. Jones. "Robust real-time face detection". *Int Journal of Computer Vision*, vol. 57, pp. 137-154, 2004.
- [56] D. G. Lowe. "Distinctive image features from scale-invariant keypoints". *Int Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

Références

- [57] T. Ahonen, A. Hadid, and M. Pietikainen. "Face recognition with local binary patterns". In ECCV, pp. 469-481, 2004.
- [58] V. Ojansivu and J. Heikkila. "Blur insensitive texture classification using local phase quantization". International Conference on Image and Signal Processing (ICISP08), pp. 236-243, 2008.
- [59] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, W. Gao, "WLD: A Robust Local Image Descriptor", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, 2010.
- [60] J. Kannala and E. Rahtu. "BSIF: Binarized Statistical Image Features". 21st International Conference on Pattern Recognition (ICPR), pp.1363-1366, 2012.
- [61] D. Lowe. "Distinctive Image Features from Scale-Invariant Key Points". Intl J. Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.
- [62] https://ena.etsmtl.ca/pluginfile.php/59680/mod_resource/content/0/SIFT%20Wikipedia.pdf.
- [63] D. Huang, Caifeng, M. Ardabilian, Y. Wang, and L. Chen. "Local Binary Patterns and Its Application to Facial Image Analysis: A Survey", IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 41, no. 6, november 2011.
- [64] T. Ojala, M. Pietikäinen, and T. Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns". IEEE Trans. Pattern Anal. Mach. Intel., vol. 24, no. 7, pp. 97-987, Jul. 2002.
- [65] H. Jin, Q. Liu, H. Lu, and X. Tong. "Face detection using improved LBP under Bayesian framework". in Proc Int. Conf. Image Graph., 2004, pp. 306-309.
- [66] H. Yang and Y. Wang. "A LBP-based face recognition method with Hamming distance constraint". in Proc. Int. Conf. Image Graph., pp. 645-649, Aug. 2007.
- [67] Z. Guo, L. Zhang, and D. Zhang. "A completed modeling of local binary pattern operator for texture classification". IEEE Trans. Image Process., vol. 19, no. 6, pp. 1657-1663, Jun. 2010.
- [68] T. Ahonen and M. Pietikäinen. "Soft histograms for local binary patterns". in Proc. Fin. Signal Process. Symp., Oulu, Finland, 2007.
- [69] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li. "Face detection based on multi-block LBP representation", Lecture Notes in Computer Science, volume 4642/2007, pages 11-18, 2007.
- [70] P. Viola et M. Jones. "Robust real-time face detection", International Conference on Computer Vision, volume 2, page 747, Juillet 2001.
- [71] Viola et M. Jones. "Rapid object detection using a boosted cascade of simple features", IEEE Conference on Computer Vision and Pattern Recognition, volume 1, pages 511–518, Décembre 2001.

Références

- [72] Y. Freund et R. Schapire. "Experiments with a new boosting algorithm", International Conference on Machine Learning, pages 148–156, Janvier 1996.
- [73] R. Lienhart et J. Maydt. "An Extended Set of Haar-like Features for Rapid Object Detection", IEEE ICIP, volume 1, pages 900-903, Septembre 2002.
- [74] M. Jones et P. Viola. "Fast Multi-view Face Detection", IEEE Conference on Computer Vision and Pattern Recognition, Juillet 2003.
- [75] T. Mita, T. Kaneko, et O. Hori. "Joint Haar-like features for face detection", IEEE International Conference on Computer Vision, volume 2, pages 1619-1626 Octobre 2005.
- [76] S. Li, L. Zhu, Z.Q. Zhang, A. Blake, H.J. Zhang, et H. Shum. "Statistical learning of multi-view face detection", European Conference on Computer Vision, Mai 2002.
- [77] G. J. Edwards, T. F. Cootes and C. J. Taylor. "Face recognition using active appearance models". In European Conference on Computer Vision, pp. 581-595, 1998.
- [78] F. Cardinaux, C. Sanderson and S. Marcel. "Comparison of MLP and GMM classifiers for face verification on XM2VTS". In International on Audio and Video Based Biometric Person Authentication, pp. 964-974, 2003.
- [79] S. Lucey and T. Chen. "A GMM parts based face representation for improved verification through relevance adaptation". In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 855-861, 2004.
- [80] F. Samaria and A. Harter. "Parameterization of a stochastic model for human face identification". In Workshop on Applications of Computer Vision, 1994.
- [81] A. V. Nefian and M. H. Hayes II I. "Face Recognition Using An Embedded HMM". In International Conference on Audio- and Video-based Biometric Person Authentication, 1999.
- [82] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade and B. Lu. "Person specific SIFT features for face recognition". In IEEE International Conference on Acoustics, Speech, and Signal Processing, 2007.
- [83] J. Ruiz del Solar, R. Verschae and M. Correa. "Recognition of faces in unconstrained environments: a comparative study". Eurasip Journal on Advances in Signal Processing, vol. 2009, no. 1, pp. 1-20, 2009.
- [84] P. J. Besl, N. D. McKay, N. (1992). "A method for registration of 3D shapes". IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 14, numéro 2. 1992.
- [85] G. Medioni and R. Waupotitsch. "Face Modeling and Recognition in 3-D". In IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp. 232-233, 2003.

Références

- [86] Y. Wang, G. Pan, Z. Wu, and Y. Wang. "Exploring facial expression effects in 3D face recognition using partial ICP". in Proc. Asian Conf. Comput. Vision (Lectures Notes in Computer Science 3851), P. Narayanan, Ed. Berlin, Germany: Springer-Verlag, pp. 581-590, Jan. 2006.
- [87] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis and T. Theoharis. "Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach". IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 4, pp. 640-649, 2007.
- [88] X. Li, T. Jia and H. Zhang. "Expression-insensitive 3D face recognition using sparse representation". In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2575-2582, Jun. 2009.
- [89] H. Mohammadzade and D. Hatzinakos. "Iterative Closest Normal Point for 3D Face Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 2, 2013.
- [90] B. Achermann, X. Jiang and H. Bunke. "Face recognition using range images". In International Conference on Virtual Systems and MultiMedia, pp. 129-136, 1997.
- [91] T. Heseltine, N. Pears and J. Austin. "Three-dimensional face recognition: an eigen surface approach". In IEEE International Conference on Image Processing, vol. 2, pp. 1421-1424, Oct. 2004.
- [92] T. Heseltine, N. Pears and J. Austin. "Three-dimensional face recognition: a fisher surface approach". In International Conference on Image Analysis and Recognition, pp. 684-691, 2004.
- [93] D. P. Huttenlocher, G. A. Klanderman and W. J. Rucklidge. "Comparing images using the Hausdorff distance". IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 9, pp. 850-863, 1993.
- [94] B. Achermann and H. Bunke. "Classifying range images of human faces with Hausdorff distance". In International Conference on Pattern Recognition, vol. 2, pp. 809-813, 2000.
- [95] G. Pan, Z. Wu and Y. Pan. "Automatic 3D face verification from range data". In IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, pp. 193-196, 2003.
- [96] T. D. Russ, K. W. Koch and C. Q. Little. "3D facial recognition : a quantitative analysis". In Annual Meeting of the Institute of Nuclear Materials Management, 2004.
- [97] X. Lu and A. K. Jain. "Deformation modeling for robust 3D face matching". In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 1377-1383, 2006.

Références

- [98] X. Lu and A. K. Jain. "Deformation modeling for robust 3D face matching". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pages 1346-1357, Aug. 2008.
- [99] X. Lu, D. Colbry and A. K. Jain. "Three-dimensional model based face recognition". In *International Conference on Pattern Recognition*, pp. 362-366, 2004.
- [100] F. R. Al-Osaimi, M. Bennamoun and A. S. Mian. "An expression deformation approach to non-rigid 3D face recognition". *International Journal of Computer Vision*, vol. 81, no. 3, pp. 302-316, 2009.
- [101] D. Huang, "Robust Face Recognition based on Three Dimensional Data", Thèse de l'école central de Lyon, 2011.
- [102] S. Gupta, M. P. Sampat and Z. Wang. "Facial range image matching using the complex wavelet structural similarity metric". In *IEEE International Workshop on Applications of Computer Vision*, page 4, Feb. 2007.
- [103] Y. Huang, Y. Wang and T. Tan. "Discriminating 3D faces by statistics of depth differences". In *Asian Conference on Computer Vision*, vol. 2, pp. 690-699, 2007.
- [104] Y. Wang, X. Tang, J. Liu, G. Pan and R. Xiao. "3D face recognition by local shape difference boosting". In *European Conference on Computer Vision*, pp. 603-616, 2008.
- [105] Y. Wang, J. Liu and X. Tang. "Robust 3D face recognition by local shape difference Boosting". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 10, pp. 1858-1870, 2010.
- [106] H. Tang, B. Yin, Y. Sun, Y. Hu. "3D face recognition using local binary patterns". *Signal Processing*, vol. 93, pp. 2190-2198, 2013.
- [107] N. Aly uz, B. Gokberk and L. Akarun. "Regional registration for expression resistant 3D face recognition". *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 425-440, 2010.
- [108] C. Samir, A. Srivastava, M. Daoudi. "Three-Dimensional Face Recognition Using Shapes of Facial Curves". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1858-1863, 2006.
- [109] L. Zhang, A. Razdan, G. E. Farin, J. Femiani, M. Bae and C. Lockwood. "3D face authentication and recognition based on bilateral symmetry analysis". *The Visual Computer*, vol. 22, no. 1, pp. 43-55, 2006.
- [110] M. H. Mahoor and M. Abdel-Mottaleb. "3D face recognition based on 3D ridge lines in range data". In *IEEE International Conference on Image Processing*, vol. 1, pp. 137-140, 2007.

Références

- [111] S. Jahanbin, H. Choi, Y. Liu and A. C. Bovik. "Three-dimensional face recognition using iso-geodesic and iso-depth curves". In IEEE International Conference on Biometrics: Theory, Applications and Systems, pp. 1-6, Oct. 2008.
- [112] H. Drira, B. Ben Amor, M. Daoudi and A. Srivastava. "Pose and expression-invariant 3D face recognition using elastic radial curves". In British Machine Vision Conference, pp. 90.1-90.11, 2010.
- [113] Z. Wu, Y. Wang and G. Pan. "3D face recognition using local shape map". In IEEE International Conference on Image Processing, volume 3, pages 2003_2006, Oct. 2004.
- [114] U. Castellani, M. Cristani, X. Lu, V. Murino and A. K. Jain. "HMM-based geometric signatures for compact 3D face representation and matching". In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Biometrics, pp. 1-6, Jun. 2008.
- [115] F. Daniyal, P. Nair and A. Cavallaro. "Compact signatures for 3D face recognition under varying expressions". In IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 302-307, Sep. 2009.
- [116] C. Maes, T. Fabry, J. Keustermans, D. Smeets, P. Suetens and D. Vandermeulen. "Feature detection on 3D face surfaces for pose normalization and recognition". In IEEE International Conference on Biometrics: Theory Applications and Systems, pp. 1-6, 2010.
- [117] S. Berretti, A. D. Bimbo, and P. Pala. "Sparse Matching of Salient Facial Curves for Recognition of 3-D Faces With Missing Parts", IEEE Transactions on Information Forensics and Security, vol.8, no. 2, 2013.
- [118] B. Gokberk, A. A. Salah and L. Akarun. "Rank-based decision fusion for 3D shape-based face recognition". In International Conference on Audio-and Video-based Biometric Person Authentication, pp. 1019-1028, 2005.
- [119] A. B. Moreno, A. Sanchez, J. Velez and J. Diaz. "Face recognition using 3D local geometrical features: PCA vs. SVM". In IEEE International Symposium on Image and Signal Processing and Analysis, pp. 185-190, Sep. 2005.
- [120] Y. Lee, H. Song, U. Yang, H. Shin and K. Sohn. "Local feature based 3D face recognition". In International Conference Audio- and Video-Based Biometric Person Authentication, pp. 909-918, 2005.
- [121] X. Li and H. Zhang. "Adapting geometric attributes for expression-invariant 3D face recognition". In IEEE International Conference on Shape Modeling and Applications, pp. 21-32, Jun. 2007.
- [122] P. Guan and L. Zhang. "3D face recognition based on facial structural angle and local region map". In IEEE International Conference on Multimedia and Expo, pp. 41-44, Apr. 2008.

Références

- [123] G. Pan, S. Han, Z. Wu and Y. Wang. "3D face recognition using mapped depth images". In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Face Recognition Grand Challenge Experiments, 2005.
- [124] G. Passalis, I. Kakadiaris, T. Theoharis, G. Toderici and N. Murtuza. "Evaluation of 3D face recognition in the presence of facial expressions: an annotated deformable model approach". In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Face Recognition Grand Challenge Experiments, pp. 171-171, 2005.
- [125] France Laliberté. "Introduction aux expressions faciales prélude au projet MONNET", Présentation, CRIM, Mai 2005.
- [126] P. Ekman et W. V. Friesen. "Facial Action Coding System (FACS): Manual", Consulting Psychologists Press, 1978.
- [127] H. Mercier. "Analyse automatique des expressions du visage : Application à la Langue des Signes", D.E.A, l'Université Paul Sabatier – Toulouse III, France, Juin 2003.
- [128] P. Ekman et W. V. Friesen. "Unmasking the Face: a guide to recognizing emotions from facial clues", livre, 2003.
- [129] I. Essa et A. Pentland. "Coding, Analysis, Interpretation, and Recognition of Facial Expressions", IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 19(7), pages 757–763, Juillet 1997.
- [130] K. Ghanem. "Reconnaissance des expressions Faciales à base d'informations Vidéo ; Estimation de l'intensité des expressions faciales", Thèse, Faculté des sciences de l'ingénieur, Algérie, Octobre 2010.
- [131] H. Maamoto. "Reconnaissance et mappage des expressions faciales vers un modèle virtuel". Thèse de Mastère 2, Département de Génie Electrique, Ecole Nationale d'ingénieur de Sousse ,2012.
- [132] M.A. Bronstein, M.M. Bronstein, R. Kimmel. "Expression-Invariant 3D Face Recognition". Proc. Audio & Video-based Biometric Person Authentication (AVBPA), Lecture Notes in Comp. Science 2688, Springer, pp. 62-69, 2003.
- [133] P. J. Phillips, P. J. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone. "Face recognition vendor test 2002: Evaluation report". NISTIR 6965. Available online at <http://www.frvt.org>, 2003
- [134] A.M. Bronstein, M.M. Bronstein, E. Gordon, R. Kimmel. "Fusion of 2D and 3D in Three-Dimensional Face Recognition". In IEEE International Conference on Image Processing, pp. 87-90, 2004.
- [135] C. Beumier, M. Acheroy. "Automatic 3D Face Authentication". Image and Vision Computing, vol. 18, no. 4, pp. 315-321. 2000.

Références

- [136] K. Chang, K. Bowyer, P. Flynn. "Face Recognition using 2D and 3D facial data". In: Proc. ACM Workshop on Multimodal User Authentication December, pp. 25–32, 2003.
- [137] Y. Wang, C. S. Chua, and Y. K. Ho. "Facial feature detection and face recognition from 2d and 3d images". Pattern Recognition Letters, vol. 23, no. 10, pp. 1191-1202, 2004.
- [138] J. Cook, C. McCool, V. Chandran, and S. Sridharan. "Combined 2D/3D Face Recognition Using Log-Gabor Templates". In Proceedings of the IEEE International Conference on Video and Signal Based Surveillance, pp. 83, 2006.
- [139] I. Mpiperis, S. Malassiotis, and M. G. Strintzis. "3-D Face Recognition With the Geodesic Polar Representation". IEEE Transactions on Information Forensics and Security, Vol. 2, No. 3, pp. 537-547, 2007.
- [140] F. R. Al-Osaimi, M. Bennamoun, and A. Mian. " Spatially Optimized Data-Level Fusion of Texture and Shape for Face Recognition". IEEE Transactions on Image Processing, Vol. 21, No. 2, 2012.
- [141] X. Wang, Q. Ruan, G. An, Y. Jin. "2D+3D Face Recognition Using Dual-tree Wavelet Transform". TENCON IEEE Region 10 Conference, pp. 1-4, 2013.
- [142] A. Mellakh. "Reconnaissance des visages en conditions dégradées". THÈSE pour l'obtention du grade de Docteur de l'Institut National des Télécommunications Spécialité : Traitement d'image et de signal, 07 Avril 2009.
- [143] C. G. Rafael and E. W Richard. "Digital Image Processing (2nd Edition) ". Prentice Hall, January 2002.
- [144] Y. Sun, X. Wang, and X. Tang. "Deep learning Face Representation from Predicting 10,000 Classes", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1891-1898.
- [145] D. Aneja, A. Colburn, G. Faigin, L. Shapiro, and B. Mones. "Modeling Stylized Character Expressions via Deep Learning", Department of Computer Science and Engineering, University of Washington, Seattle, WA, USA 2016.
- [146] D. Harwood, T. Ojala, M. Pietikäinen, S. Kelman, and S. Davis. " Texture Classification by Center-Symmetric Auto-correlation, Using Kullback Discrimination of Distributions". Technical Report, CAR-TR-678, Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park, Maryland, 1993.
- [147] T. Mäenpää, T. Ojala, M. Pietikäinen, and S. Maricor. "Robust texture classification by subsets of local binary patterns". In Pattern Recognition, International Conference on, vol 3, pp. 3947, 2000.
- [148] S. Liao and S. Z. Li. "Learning multi-scale block local binary patterns for face recognition". in Proc. Int. Conf. Biometrics, pp. 828-837, 2007.

Références

- [149] V. N. Vapnik. "The nature of statistical learning theory". Book: The nature of statistical learning theory, 1995.
- [150] L. ALLANO. "La Biométrie multimodale : stratégies de fusion de scores et mesures de dépendance appliquées aux bases de personnes virtuelles". Thèse de Docteur de l'Institut National des Télécommunications, 12 janvier 2009.
- [151] F. WANG and J. HAN. "Multimodal biometric authentication based on score level fusion using support vector machine", *Opto-Electron*. Vol. 17, no. 1, 2009.
- [152] D. A. Ramli, S. A. Samad, and A. Hussain. "Score Information Decision Fusion Using Support Vector Machine for a Correlation Filter Based Speaker Authentication System". *Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems CISIS'08*, vol. 53, pp. 235-242, 2009.
- [153] P. Ejarque, J. Hernado, D. Hernando, and D. Gómez. "Eigenfeatures and Supervectors in Feature and Score Fusion for SVM Face and Speaker Verification". *Biometric ID Management and Multimodal Communication*, vol. 5707, pp. 81-88, 2009.
- [154] D. R. Kisku, P. Gupta, and J. K. Sing. "Fusion of Multiple Matchers Using SVM for Offline Signature Identification". *Security Technology*, vol. 58, pp. 201-208, 2009.
- [155] M. Farrús, P. Ejarque, A. Temko, and J. Hernando. "Histogram Equalization in SVM Multimodal Person Verification". *Advances in Biometrics*, vol. 4642, pp. 819-827, 2007.
- [156] A. Atia. "Reconnaissance de visage pour l'authentification biométrique". Thèse de Mastère 2, Département d'informatique, Université de Mohamed Khider Biskra, Juin 2011.
- [157] W. Hwang, H. Wang, H. Kim, S. C. Kee, and J. Kim. "Face Recognition System Using Multiple Face Model of Hybrid Fourier Feature under Uncontrolled Illumination Variation". *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 1152- 1165, 2011.
- [158] A. K. Jain and A. Ross. "Multibiometric systems". *Communications of the ACM*, special issue on multimodal interfaces, vol. 47, no. 1, pp. 34-40, 2004.
- [159] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. "Impact of Artificial "Gummy" Fingers on Fingerprint Systems". In: *Proceedings of SPIE: Optical Security and Counterfeit Deterrence Techniques IV*, pp.275-289, January 2002.
- [160] I. L. Dryden, Kanti V. Mardia. "Statistical Shape Analysis", Wiley, 1998.
- [161] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski. "A comparison and evaluation of multi-view stereo reconstruction algorithms". In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 519-528, 2006.
- [162] Z. Yu, M. Y. Nam, S. Sedai, and P. K. Rhee. "Evolutionary Fusion of a Multi-Classifer System for Efficient Face Recognition". *International Journal of Control, Automation, and Systems*, vol. 7, no. 1, pp. 33-40. 2009.

Références

- [163] X. Zhang, D. Liu, and J. Chen. "An Illumination Independent Face Verification Based on Gabor Wavelet and Supported Vector Machine". *Communications in Computer and Information Science*, vol. 15, pp. 153-160, 2008.
- [164] Z. S. Zhao, J. Z. Wang, Q. J. Tian, J. M. Zhou. "Classifier Fusion of Improved Laplacian faces". *IEEE Proceedings of the Ninth International Conference on Machine Learning and Cybernetics*, Qingdao, pp. 11-14 July 2010.
- [165] Y. Wang, T. Tan, and A. Jain. "Combining face and iris biometrics for identity verification". In: *Proceedings of Fourth International Conference on Audio- and Video-Based Authentication (AVBPA)*, pp. 805-813, June 2003.
- [166] J. Kittler, M. Hatef, P. R. Duin, J. Matas. " On combining classifiers". *IEEE Trans. Pattern Analysis and Mach. Intel.* vol. 20, no. 3, pp. 226-239, 1998.
- [167] R. Brunelli and D. Falavigna. "Person identification using multiple cues". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol17, pp.955-966, 1995.
- [168] T. Ahonen, A. Hadid and M. Pietikainen. "Face Description with Local Binary Patterns: Application to Face Recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, 2006.
- [169] S. Tulyakov, S. Jaeger, and V. Govindaraju, and D. Doermann. "Review of Classifier Combination Methods". *Machine Learning in Document Analysis and Recognition*, vol. 90, pp.361- 386, 2008.
- [170] L. A. Zadeh. "Fuzzy sets". *Information Control*. vol. 8, pp. 338-353,1965.
- [171] P. Verlinde. "A contribution to multi-modal identity verification using decision fusion". PhD thesis, Ecole Nationale des Télécommunications France, 1999.
- [172] A. F. Abate, M. Nappi, D. Riccio et G. Sabatino, "2D and 3D face recognition : A survey", *Pattern Recognition Letters*, volume 28(14), pages 1885–1906, Octobre 2007.
- [173] C.L. Huang et Y.M. Huang, "Facial Expression Recognition Using Model-Based Feature Extraction and Action Parameters Classification", *Journal of Visual Communication and Image Representation*, volume 8 (3), pages 278-290, Septembre 1997.
- [174] M.J. Lyons et S. Akamatsu, "Coding facial expressions with Gabor wavelets", *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 200–205, Avril 1998.
- [175] M.J. Lyons, J. Budynek, S. Akamatsu, "Automatic Classification of Single Facial Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 21 (12), pages 1 357–1 362, Décembre 1999.
- [176] M.F. Valstar, H. Gunes et M. Pantic, "How to distinguish posed from spontaneous smiles using geometric features", *ACM International Conference on Multimodal Interfaces*, pages 38-45, 2007.

Références

- [177] S. Haykin, "Neural Networks_ A comprehensive Foundation", Livre, Macmilan College Publishing Company, 1994.
- [178] Z. Zhang, M. Lyons, M. Schuster et S. Akamatsu, "Comparison between Geometry-Based and Gabor Wavelets-Based Facial Expression Recognition Using Multi-Layer Perceptron", IEEE International Conference on Automatic Face and Gesture Recognition, pages 454–459, Avril 1998.
- [179] I. Cohen, N. Sebe, L. Chen, A. Garg et T.S. Huang. "Facial expression recognition from video sequences: temporal and static modelling", Computer Vision and Image Understanding: Special issue on face recognition, volume 91, pages 160–187, 2003.
- [180] T. Cootes, G. J. Edwards et C. J. Taylor. "Active appearance models", IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 23(6), pages 681–685, 2001.
- [181] F. Davoine, B. Abboud et M. Dang. "Analyse de visages et d'expressions faciales par modèle actif d'apparence", Traitement du Signal, Volume 21 (3), pages 179-193, février 2004.
- [182] Y. Yacoob et L.S. Davis. "Recognizing human facial expressions from long image sequences using optical flow", IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 18 (6), pages 636–642, Juin 1996.
- [183] M.J Black et Y. Yacoob. "Recognizing facial expression in image sequences using local parametrized models of image motion", International Journal on Computer Vision, volume 25 (1), pages 23–48, Octobre 1997.
- [184] Z. Hammal, A. Caplier et M. Rombaut. "A Fusion Process Based on Belief Theory Classification of Facial Basic Emotions", International Conference on Information fusion, volume 1, Juillet 2005.
- [185] F. Maalej, B. Ben Amor et M. Daoudi. "Analyse locale de la forme 3D pour la reconnaissance d'expressions faciales", ORASIS, Congrès des jeunes chercheurs en vision par ordinateur, Mai 2011.
- [186] T. Weise, S. Bouaziz, H. Li et M. Pauly. "Realtime Performance-based Facial Animation" ACM Transactions on Graphics (SIGGRAPH), volume 30(4), page 77, Juillet 2011.
- [187] V. N. Vapnik. "Statistical learning theory". Book : Statistical learning theory, 1998.
- [188] V. Brandou. "Stéréovision Locale et Reconstruction 3D/4D", Thèse, Ecole Doctorale Stic Sciences et Technologies de l'Information et de la Communication, France, Décembre 2008.
- [189] T. Chretien, M. Houdeau, O. Liandrat et F. Rixain. "Reconstruction 3D temps-réel des mains en mouvement", Projet de spécialité 2A, École Nationale Supérieure d'Informatique et de Mathématiques Appliquées de Grenoble, France, Juin 2011.
- [190] P. Sturm. "Quelques notes pour le cours de Vision par Ordinateur", INRIA Rhône-Alpes, Equipe-projet PERCEPTION.

Références

- [191] F. Cabestaing. "Reconstruction 3D géométrie épi-polaire et stéréovision", Master, Université lille1, Sciences et technologies, France.
- [192] "Larousse MEDICAL", Dictionnaire médical collectif, édition 2005.
- [193] B. Abboud, F. Davoine et M. Dang. "Facial expression recognition and synthesis based on appearance model", Signal Processing : Image Communication, volume 19(8), pages 723–740, Septembre 2004.
- [194] I. Kemelmacher-Shlizerman et R. Basri. "3D Face Reconstruction from a Single Image Using a Single Reference Face Shape". IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 33(2), pages 394-405, Février 2011.
- [195] M. Zollhöfer, M. Martinek, G. Greiner, M. Stamminger et J. Süßmuth. "Automatic Reconstruction of Personalized Avatars from 3D Face Scans". Computer Animation and Virtual Worlds, volume 22 (2-3), pages 195-202, Mars 2011.
- [196] <https://www.ri.cmu.edu/research/projects/>.
- [197] Y. Chen and G. Medioni. "Object modelling by registration of multiple range images". Image Vision Comput., vol. 10, no. 3, pp. 145-155, 1992.
- [198] S. Berretti, A. Del Bimbo, P. Pala. "Recognition of 3D Faces with Missing Parts based on Profile Networks", In Proceedings of the ACM workshop on 3D object retrieval (3DOR '10). ACM, USA, pp. 81-86, (2010).
- [199] S. Berretti, A. Del Bimbo, P. Pala. "Facial curves between keypoints for recognition of 3D faces with missing parts", IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2011, pp.46-51, 20-25 June (2011).
- [200] D. Lowe. "Demo Software: SIFT Keypoint Detector". <http://www.cs.ubc.edu.ca/lowe/>, 2006.
- [201] <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [202] <http://grail.cs.washington.edu/projects/deepexpr/ferg-db.html>.
- [203] B. Seddik, H. Maâmatou, S. Gazzah, T. Chateau, N. Ben Amara. "Unsupervised Facial Expressions Recognition and Avatar Reconstruction from Kinect", 10th International Multi-Conference on Systems, Signals & Devices (SSD) Hammamet, Tunisia, March 18-21, 2013.
- [204] <http://www.kasrl.org/jaffe.html>.
- [205] R. M. Bolle, J.H. Connell, S. Pankanti, A.W. Senior. "The Relation between the ROC Curve and the CMC", IBM Research Report, (2006).
- [206] T.C. Faltemier, K.W. Bowyer, and P.J. Flynn. "A region ensemble for 3d face recognition", IEEE Transactions on Information Forensics and Security,3(1), pp. 62-73, March (2008).

Références

- [207] N. Aly Uz, B.G. Okberk, L. Akarun. "3D face recognition system for expression and occlusion invariance", In IEEE 2nd International Conference on Biometrics: September (2008).
- [208] K.W. Bowyer, K. I. Chang and P. J. Flynn. " A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition", January (2006).