

N° d'ordre :

UNIVERSITE D'EVRY VAL D'ESSONNE

THÈSE

pour obtenir le titre de

Docteur de l'Université Evry Val d'Essonne

Spécialité:

Sciences de l'Ingénieur

Présentée par

SOUHILA GUERFI ABABSA

**Authentification d'individus par reconnaissance de
caractéristiques biométriques liées aux visages 2D/3D**

Soutenue le 03 octobre 2008 devant le jury composé de

M. M. MILGRAM, Professeur, ISIR Université de Paris VI, Président

M. L. CHEN, Professeur, EC de Lyon, Rapporteur

Mme. C. FERNANDEZ, Professeur, SIC Université de Poitiers, Rapporteur

M. A. SMOLARZ , McF HDR, LM2S Université de Troyes , Examineur

M. J-P. GAMBOTTO, PAST HDR, IBISC Université d'Evry, Co-Directeur de thèse

Mme S. LELANDAIS, Professeur, IBISC Université d'Evry, Directeur de thèse

Je dédie cette thèse

*à mes parents qui m'ont encouragé et soutenu,
à Fakhr Eddine, Selma et Sarah qui partagent ma vie,*

Remerciements

Je tiens à exprimer mes remerciements et ma vive gratitude :

A Madame Sylvie Lelandais et à Monsieur Jean-Pierre Gambotto qui m'ont permis, grâce à leur confiance et leur soutien précieux, de réaliser et surtout de mener à terme ce travail.

A Monsieur Etienne Colle, directeur du Laboratoire Informatique, Biologie Intégrative et Systèmes Complexes, pour m'avoir accueilli dans son laboratoire dans des conditions plus que favorables à une recherche fructueuse.

A Monsieur Maurice Milgram qui me fait l'honneur de présider ce jury de thèse.

A Madame Christine Fernandez et à Monsieur Liming Chen qui ont accepté d'être les rapporteurs de ce mémoire, je les remercie tout particulièrement pour l'attention et le temps qu'ils y ont consacrés.

J'exprime également toute ma sympathie et ma gratitude à tous les doctorants de l'IBISC pour l'ambiance agréable qu'ils ont su créer.

Enfin, je voudrais remercier ma famille et en particulier ma mère et mon père qui m'a donné l'éducation sans laquelle je n'en serais pas là aujourd'hui.

Table des matières

Chapitre 1 Les Systèmes Biométriques

1.1	Introduction	6
1.2	Les systèmes biométriques	8
1.3	Evaluation des performances des Systèmes biométriques	10
1.4	Fiabilité des systèmes biométriques	11
1.4.1	Test de vérification	11
1.4.2	Test d'identification.....	13
1.5	Applications des systèmes biométriques	14
1.6	La place de la reconnaissance faciale parmi les autres techniques biométriques.	15
1.7	Systèmes biométriques basés sur la reconnaissance de visage.....	16
1.7.1	Détection de visage.....	17
1.7.2	Extraction de caractéristiques du visage.....	18
1.7.3	La reconnaissance de visage.....	19
1.8	Principales difficultés de la reconnaissance de visage	19
1.8.1	Changement d'illumination	19
1.8.2	Variation de pose	20
1.8.3	Expressions faciales.....	20
1.8.4	Présence ou absence des composants structurels	21
1.8.5	Occultations partielles	21

1.9	Conclusion.....	21
-----	-----------------	----

Chapitre 2 Techniques de détection et de reconnaissance de visages

2.1	Introduction	22
2.2	Détection de visages.....	22
2.2.1	Approches basées sur les connaissances acquises.....	23
2.2.2	Approches basées sur le « Template-matching »	24
2.2.3	Approches basées sur l'apparence.....	25
2.2.4	Approches basées sur des caractéristiques invariantes.....	27
2.2.5	Détection de visages basée sur l'analyse de la couleur de la peau.....	29
2.2.6	Comparaison des différentes approches	36
2.3	Techniques 2D de reconnaissance de visage.....	37
2.3.1	Approches holistiques ou globales	38
2.3.2	Méthodes locales	44
2.3.3	Méthodes Hybrides.....	52
2.4	Techniques 3D de reconnaissance de visages	55
2.4.1	Systèmes d'acquisition 3D	55
2.4.2	Approches modèle.....	57
2.4.3	Approches 3D.....	57
2.4.4	Approches 3D+2D.....	61
2.4.5	Conclusion.....	63

Chapitre 3 Bases de données utilisées

3.1	Généralité sur les bases de données existantes.....	66
3.2	La base XM2VTS [Mes99]	68
3.3	La base IV ² [IV05]	69
3.3.1	Protocoles d'acquisition IV ²	70

3.3.2	Evaluation préliminaire	71
3.3.3	Protocoles	73
3.3.4	Conclusion	74

Chapitre 4 Extraction de visage et de ses caractéristiques

4.1	Introduction	75
4.2	Segmentation couleur du visage	76
4.3	Segmentation par Ligne de Partage des Eaux	77
4.3.1	Principe de l'immersion	77
4.3.2	Ligne de Partage des Eaux (d'après Vincent et Soille)	81
4.4	Approche segmentation de couleur proposée	82
4.4.1	Représentation de l'espace TLS	82
4.4.2	Détermination de l'image de la norme du gradient TLS	84
4.4.3	Simulation de l'inondation d'après Vincent	88
4.4.4	Fusion des bassins versants	91
4.4.5	Extraction de visage	98
4.4.6	Extraction des régions caractéristiques du visage	99
4.4.7	Extraction et Normalisation des régions pour la vérification de visage	109
4.4.8	Conclusion	110

Chapitre 5 Reconnaissance de visage 2D/3D

5.1	Introduction	111
5.2	Authentification de visage 2D	112
5.2.1	Eigenface globale	113
5.2.2	Eigenface modulaire	115
5.2.3	Tests et Evaluations	116
5.3	Vérification de visage 3D	125

5.3.1	Etude de l'anthropométrie du visage	126
5.3.2	Sélection des points caractéristiques	129
5.3.3	La signature 3D	130
5.3.4	Système d'acquisition de visage 3D.....	132
5.3.5	Protocole d'évaluation.....	137
5.4	Vérifications de visage 2D/3D	141
5.4.1	Evaluation de la méthode proposée.....	143
5.4.2	Comparaison des approches 2D, 3D et 2D/ 3D.....	143
5.4.3	Identification visage	145
5.5	Discussion et Conclusion	147
Conclusion et Perspectives		149
Annexe A		
Algorithmes.....		151
Annexe B		
	Calibration de la Caméra	155
B.1	La projection perspective	156
B.2	Transformation caméra/image.....	156
B.3	Les paramètres intrinsèques	156
B.3	Les paramètres extrinsèques.....	157
B.4	La transformation mire/image	158
B.5	Estimation des coefficients de la matrice M.....	159
Annexe C		
Indices anthropométriques.....		162
Références		163

Liste des figures

Figure 1. 1. Caractéristiques biométriques.	7
Figure 1.2. Principaux modules d'un système biométrique ainsi que les différentes modes.	9
Figure 1.3. Courbe du point d'équivalence des erreurs dans un système biométrique.	12
Figure 1. 4. Courbe DET.	13
Figure 1. 5. Courbe CMC.	14
Figure 1. 6. Les étapes de la reconnaissance de visage.	17
Figure 1. 7. Exemple de variation d'éclairage.	20
Figure 1. 8. Exemples de variation de poses.	20
Figure 1. 9. Exemples de variation d'expressions.	21
Figure 2. 1 Modèle de visage composé de 16 régions (les rectangles) associées à 23 relations (flèches). [Sin94]	24
Figure 2. 2. Différentes régions utilisées pour la phase de template matching	25
Figure 2.3. Les dix vues d'une personne dans la base de données ORL.	39
Figure 2.4. Taux d'identification moyen en fonction du nombre d'exemples d'apprentissage par personne [Wan03].	40
Figure 2. 5. (a) image originale (b) carte de projection ; (c) image combinée.	42
Figure 2. 6. Synthèse de nouvelles images avec différentes dégradations [Jun04].....	43
Figure 2.7. Image de visage partitionnée en imagettes[Che04].	48

Figure 2.8. Exemple : (a) une image du visage originale, (b) sa projection (SOM-Face) et (c) image reconstruite [Tan05].	50
Figure 2. 9. Processus de reconnaissance de visages basé sur les Modèles Actifs d'Apparence [Lan95].	53
Figure 2. 10. (a) Image texture (b) Image 2.5D (c) Image 3D	56
Figure 2. 11. Choix de dix points : points 2D marqués par le " " et points 3D marqués par "×" [Wan02].	62
Figure 3. 1. Exemple d'images extraites de la base XM2VTS.	65
Figure 3. 2. Exemple d'extraction de 10 images d'une personne dans al base de développement.	68
Figure 3.3. Normalisation des visages présentés (figure 3.1).	69
Figure 4. 1. Minima, bassins versants et ligne de partage des eaux.	78
Figure 4. 2. Le plus court chemin géodésique entre s et t .	78
Figure 4. 3. Exemple de SKIZ géodésique.	79
Figure 4. 4. Les différents cas de l'immersion.	80
Figure 4. 5. Représentation numérique circulaire de la teinte	84
Figure 4. 6. Modélisation du gradient multi composantes avec l'approche Di Zenzo.	85
Figure 4. 7. Mesure de la pertinence de la teinte	87
Figure 4. 8. Les différentes composantes : teinte (b), luminance (c), saturation (d), gradient(e).	88
Figure 4. 9. Les différentes composantes : teinte (b), luminance (c), saturation (d), gradient(e).	88
Figure 4. 10. Images originales (a) et résultats de l'immersion (b).	90
Figure 4. 11. Principe de la fusion des bassins versants.	92
Figure 4. 12. Stratégie d'utilisation des composantes TLS.	93
Figure 4. 13. Exemples de segmentation de visage : (a) images originales (b) images LPE modifié.	95

Figure 4. 14. (a) image originale (b) image de saturation (c) histogrammes de la saturation et seuils obtenus avec la méthode de Cheng.....	97
Figure 4. 15. (a) image originale (b) image de saturation (c) histogrammes de la saturation et seuils obtenus avec la méthode de Cheng.....	98
Figure 4. 16. (a) image originale, (b) résultats d'extraction du visage.....	99
Figure 4. 17. Modèle géométrique du visage.	101
Figure 4. 18. (a) image originale, (b) visage segmenté I_s , (c) visage binarisé I_M , (d) les centres de régions yeux classifiées.	102
Figure 4. 19. Processus de fusion des régions appartenant à l'œil droit.....	103
Figure 4. 20. Résultats d'extraction des régions caractéristiques.....	104
Figure 4. 21. (a) image originale, (b) mask Visage (c) régions extraites	105
Figure 4. 22. (a) image originale (b) mask visage (c) régions extraites	106
Figure 4. 23. (a) les bonnes détections (b) les mauvaises détections.	108
Figure 4. 24. (a,b) les bonnes détections (c) les mauvaises détections.....	108
Figure 4. 25. (a) Les caractéristiques extraites avec notre méthode (b) visage normalisé (c) les régions extraites.	109
Figure 5. 1. Exemple de décomposition en Eigenfaces.....	114
Figure 5. 2. Courbes DET : les résultats sur la deuxième expérience.	118
Figure 5. 3. Régions faciales utilisées pour construire les quatre classifieurs.....	119
Figure 5. 4. Courbe DET : des différents classifieurs.	120
Figure 5. 5. Fusion des scores. La classification se fait grâce à des règles permettant de combiner les sorties des quatre classifieurs.	121
Figure 5. 6. Les courbes DET des différentes combinaisons de classifieurs.....	122
Figure 5. 7. Courbes DET : régionM extraction manuelle, régionA extraction automatique.	124
Figure 5. 8. Les différents résultats de fusion de classifieurs.....	125

Figure 5. 9. Dessin au dioptrigraphe du Crâne1 montrant les points et paramètres utilisés dans la craniophotocomparaison	128
Figure 5. 10. Points (a), paramètres (b) et angles (c) utilisés pour la création d'une signature 3D facial.....	131
Figure 5. 11. Système stéréoscopique utilisé.....	133
Figure 5. 12. Mire de calibration	134
Figure 5. 13. Différents points choisis pour la reconstruction.....	134
Figure 5. 14. Vision stéréoscopique.	135
Figure 5. 15. Images stéréo extraites de la base d'évaluation IV2 : a) image droite, b) image gauche.....	138
Figure 5. 16. Courbes DET : Comparaison entre les approches, 3D et visage 2D global..	139
Figure 5. 17 Analyse des comparaisons des comparaisons inter et intra classe.....	139
Figure 5. 18. Exemple des 4 données choisies pour la reconnaissance du visage	142
Figure 5. 19. Schéma de principe de l'approche multi modale 2D-3D	143
Figure 5. 20. Exemples d'images utilisées pour l'évaluation de l'approche 2D-3D.....	143
Figure 5. 21. Courbes DET : Comparaison entre les approches, 2D modulaire, 2D globale et 3D	144
Figure 5. 22. Les Taux ERR des différentes approches.	145
Figure 5. 23. Comparaison des courbes CMC de 2D, 3D et 2D/3D, sur la sous-base IV2 . Un visage est reconnu au rang r si une vue du même visage est parmi ses r plus proches voisins.....	146
Figure 5. 24. Comparaison des courbes CMC 2D/3D (Vote à la majorité) et 2D/3D (Somme pondérée), sur la sous-base IV2 phase 2.	147

Introduction Générale

De nos jours on parle de plus en plus de l'insécurité dans divers secteurs ainsi que des moyens informatiques à mettre en œuvre pour contrer cette tendance : le contrôle d'accès aux ordinateurs, l'e-commerce, les opérations bancaires basés sur l'identification du demandeur, etc. Il existe traditionnellement deux manières d'identifier un individu. La première méthode est basée sur une connaissance à priori "knowledge-based" de la personne telle que, par exemple, la connaissance de son code PIN qui permet d'activer un téléphone portable. La seconde méthode est basée sur la *possession* d'un objet "token-based". Il peut s'agir d'une pièce d'identité, d'une clef, d'un badge, etc. Ces deux modes d'identification peuvent être utilisés de manière complémentaire afin d'obtenir une sécurité accrue comme dans le cas de la carte bleue. Cependant, elles ont chacune leurs faiblesses. Dans le premier cas, le mot de passe peut être oublié par son utilisateur ou bien deviné par une autre personne. De plus une personne sur quatre seulement fait l'effort d'appliquer les consignes de sécurité avant de retirer de l'argent (regarder derrière soi, cacher le clavier avec sa main lors de la saisie du code secret, etc.). Dans le second cas, le badge (ou la pièce d'identité ou la clef) peut être perdu ou volé. Les caractéristiques *biométriques* sont une solution alternative aux deux modes d'identification précédents. L'avantage de ces caractéristiques biométriques est d'être *universelles*, c'est-à-dire présentes chez toutes les personnes à identifier. D'autre part, elles sont *mesurables* et *uniques* : deux personnes ne peuvent posséder exactement la même caractéristique. Elles sont aussi permanentes ce qui signifie qu'elles ne varient pas ou peu au cours du temps. L'intérêt des applications utilisant la biométrie se résume en deux classes : faciliter le mode de vie, éviter la fraude.

La reconnaissance faciale fait partie des techniques biométriques. On remarque que dans la vie quotidienne chacun de nous identifie tout au long de la journée différents visages. Ainsi lorsque nous rencontrons une personne, notre cerveau va chercher dans notre mémoire et vérifier si cette personne est répertoriée ou non. La difficulté de la reconnaissance de visage par ordinateur varie énormément suivant que les conditions d'acquisition. 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 faire la reconnaissance à proprement parler. Il faut tout d'abord *détecter* la présence ou l'absence de visage dans l'image. Le visage doit ensuite être segmenté. Enfin, si nous travaillons sur un flux vidéo, le système doit suivre le visage d'une image à la suivante.

Plusieurs méthodes ont été développées pour la reconnaissance de visage 2D. Cependant, elles présentent un certain nombre de limitations liées à l'orientation du visage ou à la pose, à l'éclairage, à l'expression faciale, aux occultations, etc. 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 cette thèse nous proposons de développer une technique bimodale 2D-3D de reconnaissance faciale qui combine des mesures anthropologiques 3D du visage avec une technique 2D basée sur l'Analyse en Composantes Principales ou ACP. L'intérêt d'une telle approche est d'exploiter les avantages des deux techniques (2D et 3D) afin d'améliorer le taux de reconnaissance.

Notre but étant de développer un système d'authentification de visage simple et efficace dans le cadre du projet IV², il est nécessaire de travailler à tous les niveaux du système (détection, extraction des caractéristiques et reconnaissance) en apportant des contributions pertinentes à différents points de la chaîne de traitement. Ainsi, nous avons commencé par développer une technique « bas niveau » pour la localisation de visage unique sur fond uniforme par un procédé simple, et le plus robuste possible aux variations d'éclairage, d'où l'exploitation de techniques de traitements d'images couleur classiques ou il s'agit de travailler sur une segmentation couleur robuste. Cette approche est basée sur

un algorithme efficace de segmentation couleur du visage dans l'espace TLS (Teinte, Luminance, Saturation) qui utilise la méthode de Lignes de Partage des Eaux (LPE). Ce choix se justifie par le fait que la méthode LPE est très efficace pour segmenter des régions convexes et fermées, ce qui est le cas du visage humain. Par ailleurs, nous proposons de modifier l'algorithme LPE classique en ajoutant un critère basé sur la teinte pour la fusion des bassins versants.

L'extraction de la région du visage à partir de l'image segmentée est une étape primordiale dans le processus de reconnaissance faciale. L'algorithme d'extraction doit être efficace et doit fournir, sans aucune ambiguïté, la région du visage dans l'image. Nous avons proposé une méthode originale basée sur la classification de la couleur de la peau, pour extraire la région du visage à partir de l'image segmentée. Pour cela, nous avons établi deux règles empiriques : la première est basée sur une méthode explicite qui définit un domaine d'appartenance de la couleur de la peau dans le plan Teinte. La deuxième règle définit la taille minimale des régions considérées. Par ailleurs, nous avons défini un modèle géométrique pour extraire les régions caractéristiques du visage. Ce modèle est basé sur les distances « interoculaires » et « yeux-bouche » et permet de classifier trois régions pertinentes du visage, à savoir la région « œil droit », la région « œil gauche » et la région « bouche ».

Une autre contribution de cette thèse a consisté à développer une technique 2D de reconnaissance du visage basée sur l'Analyse en Composante Principale qui prend en entrée, non pas l'image entière du visage, mais les « imagettes » correspondant aux trois régions caractéristiques du visage (les yeux, le nez et la bouche) et à démontrer que cette méthode donne des taux de reconnaissance aussi bons que l'image complète.

D'autre part, nous pensons que l'anthropométrie par sa maîtrise de l'anatomie du visage humain peut apporter des informations complémentaires pour améliorer la reconnaissance faciale. En effet, les anthropologues ont été les premiers à étudier de façon scientifique la morphologie humaine en général et celle de la tête et du visage en particulier et cela en utilisant des critères descriptifs et des caractères métriques. Nous proposons de définir des mesures anthropométriques faciales à partir de points caractéristiques 3D du visage, lesquels sont reconstruits par stéréovision. Ces mesures correspondent à des valeurs indiciaires et angulaires caractérisant les différentes régions du visage. Une méthode

algébrique qui exploite ces mesures anthropologiques a été ensuite développée afin de réaliser la reconnaissance d'individus. Les différents tests réalisés sur la base IV^2 ont permis de valider cette approche.

Enfin, nous proposons de fusionner les résultats de reconnaissance obtenus par les deux approches (2D et 3D) afin d'améliorer le taux de reconnaissance. Pour cela, nous avons exploré plusieurs critères de fusion. Un protocole expérimental a été mis en oeuvre pour tester cette méthode et analyser ses performances.

Organisation de la thèse

Nous avons choisi d'articuler notre étude autour de cinq chapitres principaux.

Le premier chapitre est consacré à la présentation générale de la biométrie. Il décrit tout d'abord le principe de fonctionnement des systèmes biométriques puis définit les outils utilisés pour évaluer leurs performances. Ensuite, la place de la reconnaissance faciale parmi les autres techniques biométriques est analysée. A travers ce chapitre, nous voulons positionner le problème de la reconnaissance faciale et présenter ses enjeux et intérêts par rapport aux autres techniques. Nous étudions également les principales composantes d'un système de reconnaissance faciale, notamment la détection, l'extraction de caractéristiques et la reconnaissance. Enfin, nous mettons en lumière les difficultés rencontrées par les systèmes de reconnaissance de visage que nous proposons d'étudier dans le cadre de cette thèse.

Dans le second chapitre une analyse détaillée des différentes techniques développées au cours de ces dernières années dans les domaines de la détection de visage et de la reconnaissance 2D et 3D est présentée. Notre objectif est de classifier ces méthodes de manière efficace afin de mettre en évidence les particularités ainsi que les avantages et les inconvénients de chacune d'entre elles. Ceci nous permettra, par la suite, de mettre en avant les critères de performances qui vont guider le choix des solutions retenues dans le cadre de notre problématique.

Dans le troisième chapitre nous présentons les deux bases données utilisées dans le cadre de cette thèse, XM2VTS, et IV^2 . Nous détaillerons particulièrement la phase de développement de la base IV^2 dans laquelle nous avons été impliquée.

Le quatrième chapitre est dédié à notre approche d'extraction de visage et de ses caractéristiques. Nous présentons, dans un premier temps, le problème de la segmentation de visage dans une image couleur. Ensuite, nous détaillons la méthode de segmentation que nous avons élaborée, basée sur l'algorithme de la Ligne de Partage des Eaux (LPE). Par ailleurs, nous décrivons les différents outils que nous avons mis en place afin d'implémenter notre méthode de segmentation, notamment l'espace TLS et le gradient vectoriel TLS. Puis, nous détaillons les quatre étapes de notre algorithme d'extraction, à savoir : création des supports de régions, fusion de régions, extraction de visage et extractions des régions caractéristiques. Les résultats obtenus à chaque étape sont examinés et des conclusions en sont tirées.

Dans le cinquième chapitre, nous nous intéressons à l'exploitation des informations 2D et 3D du visage comme moyen biométrique pour l'authentification de personnes. Dans un premier temps, nous décrivons le dispositif d'acquisition que nous avons choisi. Il se compose d'une paire stéréoscopique pour l'acquisition d'images 2D et la reconstruction de données 3D. Ensuite, nous détaillons les différentes approches de reconnaissance faciale que nous avons élaborées, à savoir : l'approche 2D basée ACP, l'approche 2D modulaire, l'approche 3D et l'approche bimodale 2D-3D. Nous présentons les résultats expérimentaux obtenus par chaque méthode et analysons les performances.

Enfin, la conclusion générale résumera nos contributions et donnera quelques perspectives sur les travaux futurs.

Chapitre 1

Les Systèmes Biométriques

1.1 Introduction

La *biométrie* consiste à identifier une personne à partir d'une ou de plusieurs caractéristiques physiologiques (empreintes digitales, visage, iris, contour de la main, etc.), ou comportementales (signature, démarche, etc.). Etymologiquement, la biométrie humaine est synonyme d'anthropologie physique [Oli61]. Une autre définition de la *biométrie* est donnée par Roethenbaugh [Roe98]: «La biométrie s'applique à des particularités ou des caractères humains uniques en leur genre et mesurables, permettant de reconnaître ou de vérifier automatiquement l'identité ».

Les systèmes biométriques sont de plus en plus utilisés depuis quelques années. L'apparition de l'ordinateur et sa capacité à traiter et à stocker les données ont permis la création des systèmes biométriques informatisés. Il existe plusieurs caractéristiques physiques uniques pour un individu, ce qui explique la diversité des systèmes appliquant la biométrie, selon que l'on prend en compte :

- L'empreinte digitale
- La géométrie de la main
- L'iris
- La rétine

- Le visage
- La dynamique de frappe au clavier
- La reconnaissance vocale
- La dynamique des signatures
- L'ADN
-

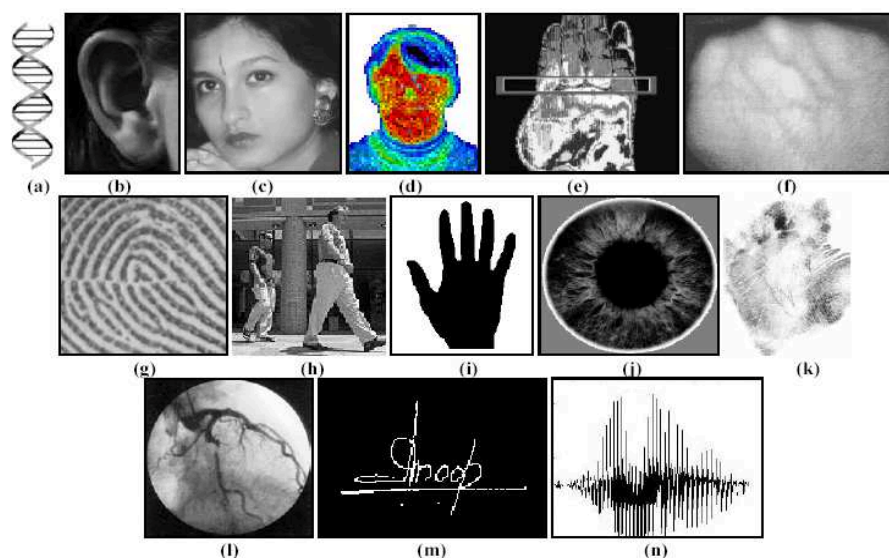


Figure 1. 1. Caractéristiques biométriques: a) ADN, b) Oreille, c) visage, d) visage infrarouge, e) thermogramme main, f) veine main, g) Empreintes digitales, h) marche, i) geste j) iris, k) empreinte de la paume, l) rétine, m) signature, n) voix.

Nous introduirons dans ce chapitre quelques notions et définitions de base liées à la biométrie. Nous donnerons le principe de fonctionnement des systèmes biométriques ainsi que les outils utilisés pour mesurer leurs performances. Nous insisterons surtout sur la place de la reconnaissance faciale parmi les autres techniques biométriques, car elle constitue l'objectif de cette thèse. Enfin, nous exposerons les difficultés majeures liées à la reconnaissance de visage, et qui font toujours l'objet de recherche par la communauté scientifique.

1.2 Les systèmes biométriques

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 :

- Le mode d'**enrôlement** est une phase d'apprentissage qui a pour but de recueillir des informations biométriques sur les personnes à identifier. Plusieurs campagnes d'acquisitions de données peuvent être réalisées afin d'assurer une certaine robustesse au système de reconnaissance aux variations temporelles des données. Pendant 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. Le traitement lié à l'enrôlement n'a pas de contrainte de temps, puisqu'il s'effectue « hors-ligne ».
- Le mode de **vérification ou authentification** est une comparaison "un à un", dans lequel le système valide l'identité d'une personne en comparant les données biométriques saisie 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 ?*». Actuellement la vérification est réalisée via un numéro d'identification personnel, un nom d'utilisateur, ou bien une carte à puce.
- Le mode d'**identification** est une comparaison "un 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 ?* ».

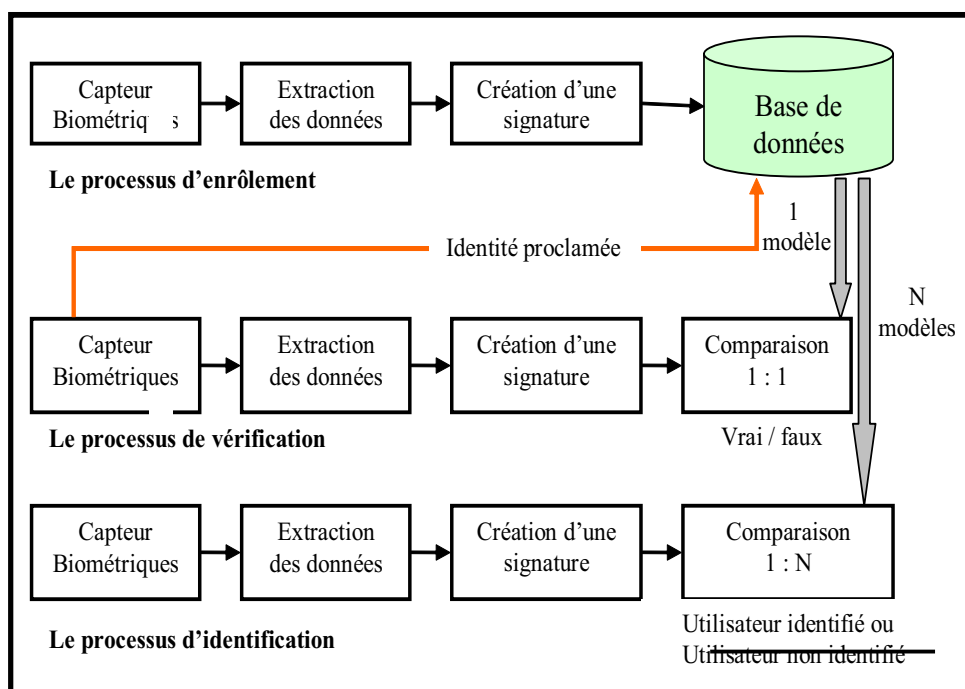


Figure 1.2. Principaux modules d'un système biométrique ainsi que les différents modes.

Les différents modules qui composent un système biométrique sont représentés sur la figure 1.2 ; leur fonctionnement peut être résumé comme suit :

- *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 de visage ou des régions caractéristiques de 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 à contrôle (volontairement ou à son insu) avec les « signatures » mémorisées. Ce module fonctionne soit en mode vérification (pour une identité proclamée) ou bien en mode identification (pour une identité recherchée).
- *Module base de données* : stocke les modèles biométriques des utilisateurs enrôlés.

1.3 Evaluation des performances des Systèmes biométriques

Chaque caractéristique (ou modalité) biométrique a ses forces et ses faiblesses, et le choix dépend de l'application visée. On ne s'attend à ce qu'aucune modalité biométrique ne réponde efficacement aux exigences de toutes les applications. En d'autres termes, aucun système biométrique n'est "optimal". Faire correspondre un système biométrique spécifique à une application dépend du mode opérationnel de l'application et des caractéristiques biométriques choisies. Plusieurs études ont été menées afin d'évaluer les performances des systèmes biométriques. La société américaine – l'*International Biometric Group* [IBG] – à par exemple effectué une étude basée sur quatre critères d'évaluation :

- *intrusivité*: ce critère permet de classifier les systèmes biométriques en fonction de l'existence d'un contact direct entre le capteur utilisé et l'individu à reconnaître. La reconnaissance faciale est une technique « non intrusive », car il n'existe aucun contact entre le capteur (la caméra) et le sujet, elle est bien acceptée par les utilisateurs à l'inverse d'autres techniques « intrusives » comme l'iris où un contact direct est nécessaire entre le capteur et l'œil.
- *fiabilité* : dépend de la qualité de l'environnement (éclairage par exemple) dans lequel l'utilisateur se trouve. Ce critère influe sur la reconnaissance de l'utilisateur par le système. Nous verrons ce point en détail dans la section suivante.
- *coût* : doit être modéré. À cet égard nous pouvons dire que la reconnaissance faciale ne nécessite pas une technologie coûteuse. En effet, la plupart des systèmes fonctionnent en utilisant un appareil à photo numérique de qualité standard.
- *effort* : requis par l'utilisateur lors de la saisie de mesures biométriques, et qui doit être réduit le plus possible. La reconnaissance faciale est la technique biométrique la plus facile à utiliser car non contraignante.

Les résultats de cette étude peuvent être présentés comme suit :

- les techniques les moins "intrusives" aux plus "intrusives" : la voix, la frappe sur le clavier, la signature, la main, le visage, l'empreinte digitale, l'iris et enfin la rétine.

- les techniques les plus fiables aux moins fiables: l'iris, la rétine, l'empreinte digitale, le visage, la main, la voix, et enfin à un niveau équivalent, la frappe sur le clavier et la signature.
- les techniques les moins coûteuses aux plus coûteuses: la frappe sur le clavier, la voix, la signature, l'empreinte digitale, le visage, la main, la rétine et enfin l'iris.
- les techniques les plus faciles d'utilisation aux plus difficiles: le visage, la signature, l'iris, la frappe sur le clavier, la voix, l'empreinte digitale, la main et enfin la rétine.

1.4 Fiabilité des systèmes biométriques

Afin de mesurer la fiabilité d'un système biométrique en modes de vérification et/ou d'identification, deux principaux tests sont utilisés :

1.4.1 Test de vérification

Dans la tâche de vérification, un utilisateur final doit faire une demande d'authentification de son identité. Par exemple : il proclame "je suis Mr Dupont", alors le système biométrique doit déterminer si l'identité proclamée par l'utilisateur est acceptée ou rejetée. Deux taux sont alors calculés :

- *Le Taux de Faux Rejets* ou False-Rejection Rate (FRR), il exprime le pourcentage d'utilisateurs rejetés alors qu'ils devraient être acceptés par le système.
- *Le Taux de Fausses Acceptations* ou False-Acceptance Rate (FAR), il exprime le pourcentage d'utilisateurs acceptés par le système alors qu'ils devraient être rejetés.

On peut formuler le test de vérification de la façon suivante [Jai04] :

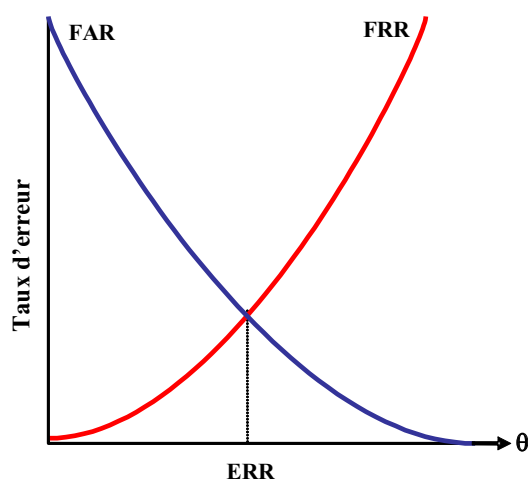
Soient X_Q le vecteur de caractéristiques de la personne proclamée I , X_I le vecteur de caractéristiques de la personne I stockée dans la base de données, $S(X_Q, X_I)$ la fonction de similarité entre le vecteur X_Q et X_I . La fonction S donne le score de similarité entre les mesures biométriques de la personne de la base de données et la personne proclamée.

Le test de vérification est alors défini par la fonction (I, X_Q) , telle que :

$$(I, X_Q) = \begin{cases} w_1 & \text{Si } S(X_Q, X_I) \geq \theta \\ w_2 & \text{Sinon} \end{cases} \quad (1.1)$$

Où w_1 indique que la personne proclamée est vraie et w_2 qu'elle est un imposteur.

Le choix du seuil de similarité θ est important car il influe directement sur les performances du système. Un seuil θ trop petit entraîne l'apparition d'un grand nombre de faux rejets, tandis qu'un seuil θ trop grand engendre un taux important de fausses acceptations. La statistique la plus simple pour mesurer la performance d'un algorithme dans le contexte de la vérification est de calculer le *point d'équivalence des erreurs* (*Equal Error Rate - EER*). Le point d'équivalence des erreurs, ou taux d'exactitude croisée, est déterminé par le point d'intersection entre la courbe du taux de fausses acceptations et la courbe du taux de faux rejets. Un exemple de courbes d'erreurs croisées est donné à la figure (1.3).



Source: Mansfield, T. *et al.* (2001),
"Biometric Product Testing Final Report"

Figure 1.4. Courbe du point d'équivalence des erreurs dans un système biométrique.

L'évaluation passe également par le tracé de statistiques complexes, comme la courbe « Detection Error trade-off (DET) ». Cette courbe donne le FRR en fonction du FAR. Elle est tracée de manière paramétrique en fonction des valeurs du seuil θ .

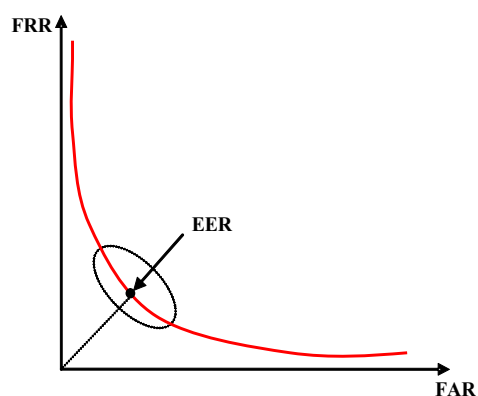


Figure 1. 5. Courbe DET.

1.4.2 Test d'identification

On peut formuler le test d'identification de la façon suivante:

Soient X_Q le vecteur de caractéristiques, pour déterminer l'identité I_k , $k \in \{1, 2, \dots, N, N+1\}$, on calcule la fonction (I, X_Q) définie par :

$$(I, X_Q) = \begin{cases} I_k & \text{Si } \max_k \{S(X_Q, X_{I_k})\} \geq \theta, k = 1, \dots, N, \\ I_{N+1} & \text{Sinon} \end{cases} \quad (1.2)$$

où I_1, \dots, I_N sont les identités enrôlées, I_{N+1} une identité rejetée, X_{I_k} la signature biométriques qui correspond à l'identité I_k , et θ le seuil.

Le *test d'identification* représente la mesure la plus couramment utilisée, mais il n'est pas toujours suffisant. En effet, en cas d'erreur, il peut être utile de savoir si le bon choix se trouve parmi les N premières réponses du système. On trace alors la courbe *Cumulative Match Characteristics* (CMC) qui représente la probabilité que le bon choix se trouve parmi les N premiers [Phi00]. Comme l'illustre la figure 1.5.

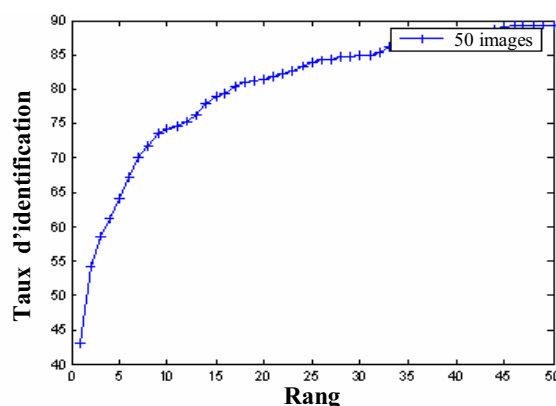


Figure 1. 6. Courbe CMC.

1.5 Applications des systèmes biométriques

Les applications de la biométrie peuvent être divisées en trois groupes principaux:

- *Applications commerciales* : telles que l'ouverture de réseau informatique, la sécurité de données électroniques, l'e-commerce, l'accès Internet, la carte de crédit, le contrôle d'accès physique, le téléphone cellulaire, la gestion des registres médicaux, l'étude à distance, etc.
- *Applications gouvernementales* : telles que la carte d'identité nationale, le permis de conduire, la sécurité sociale, le contrôle des frontières, le contrôle des passeports, etc.
- *Applications légales* : telles que l'identification de corps, la recherche criminelle, l'identification de terroriste, etc.

De nos jours les systèmes biométriques sont de plus en plus utilisés dans des applications civiles. Par exemple, le dispositif de Schiphol Privium à l'aéroport d'Amsterdam, utilise un capteur de l'iris pour accélérer la procédure de contrôle des passeports et des visas [CNN02]. Les passagers insèrent leur carte dans un lecteur et se mettent en face d'un appareil-photo ; ce dernier acquiert l'image de l'œil. Des processus de traitement d'images sont alors lancés afin de localiser l'iris et de calculer une signature appelée « Iriscode » [Dau99]. Une fois l'Iriscode calculé, il est comparé aux données stockées dans la carte pour identifier le passager. Un dispositif semblable est également employé pour vérifier l'identité des employés de l'aéroport qui travaillent dans des secteurs de haute sécurité. Un autre exemple d'utilisation à grande échelle de la reconnaissance faciale est celui de la ville

de Newham, dans la banlieue de Londres. Il s'est vu décerner le trophée *Big Brother Award* en 1998 par l'organisation non gouvernementale *Privacy International*.

A en croire les perspectives de l'*International Biometric Group*, le marché de la sécurité des biens et des personnes ne s'est jamais aussi bien porté. En 2005, le marché représentait environ 1,8 milliard de dollars et pourrait atteindre les 5 milliards en 2008. Les technologies les plus répandues sont les empreintes digitales (48% du marché), la morphologie faciale (15,4% du marché), la géométrie de la main (11%), l'iris (9%), la reconnaissance vocale (6%), la rétine, la morphologie des oreilles ainsi que la radiographie dentaire. Notons que les techniques comportementales (comme la démarche, le sourire, le mouvement des yeux, etc.) ont quand à elles du mal à s'imposer.

1.6 La place de la reconnaissance faciale parmi les autres techniques biométriques

Les empreintes digitales sont les caractéristiques biométriques les plus communément utilisées pour la reconnaissance de criminels. Le premier système automatique d'authentification utilisant les empreintes digitales a été commercialisé au début des années soixante. D'autres parts, plusieurs études ont démontré que l'iris est la caractéristique la plus fiable car la texture de l'iris reste stable au cours de la vie (c.f. section 1.3). Toutefois, ces méthodes présentent l'inconvénient majeur d'être intrusives, ce qui limite énormément leurs domaines d'applications. De plus, une méthode comme l'identification de l'iris reste contraignante pour les utilisateurs qui n'apprécient pas de placer leur œil devant un appareil. A l'inverse, des systèmes d'identification basés sur l'analyse des images de 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 d'améliorer les performances de la reconnaissance.

Depuis quelques années, 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 [Phi98b, 00; Riz98], XM2VTS [Mes99], FRVT 2000 [Bla01] et FRVT 2002 [Phi03].

Dans [Pra07] les auteurs présentent une analyse statistique des publications sur les techniques biométriques soumises et publiées dans un numéro spécial de la revue *IEEE Transaction on PAMI* (voir tableau 1.1). Nous constatons que la reconnaissance faciale arrive largement en tête avec un pourcentage de 33% du nombre total de publications. Ceci démontre bien l'intérêt scientifique pour cette technique.

Article (%)	Visage	Empreintes digitale	Multimodale	Iris	Performance Evaluation	Autres
Soumission	33%	17%	16%	9%	4%	21%
Acceptation	33%	16%	16%	11%	5%	20%

Tableau 1. 1. Répartition des articles sur les techniques biométriques soumis et acceptés dans la revue IEEE PAMI.

1.7 Systèmes biométriques basés sur la reconnaissance de visage

La reconnaissance automatique de visage s'effectue en trois étapes principales : (1) détection de visages, (2) extraction et normalisation des caractéristiques du visage, (3) identification et/ou vérification (voir figure 1.6). Certaines techniques de traitements d'images peuvent être communes à plusieurs étapes. Par exemple, l'extraction des caractéristiques faciales (yeux, nez, bouche) est utilisée aussi bien pour la détection que pour l'identification de visages. Par ailleurs, les étapes de détection de visage et d'extraction de caractéristiques peuvent être exécutées simultanément. Cela dépend notamment de la nature de l'application, de la taille de la base d'apprentissage, et des conditions de prise de vue (bruit, occultation, etc.). Enfin, les techniques de traitement utilisées dans chaque étape sont très critiques pour les applications biométriques, et

doivent, par conséquence, être optimisées pour améliorer les performances du système global.

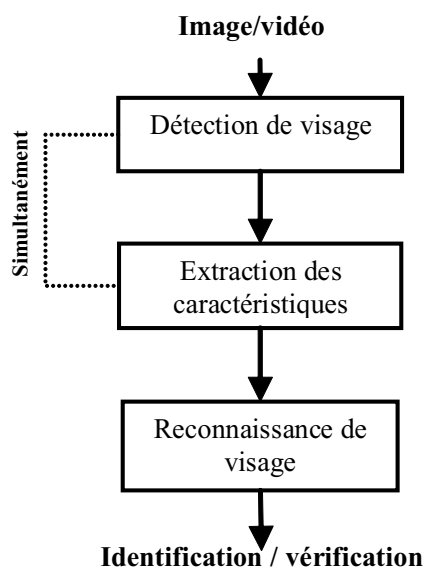


Figure 1. 7. Les étapes de la reconnaissance de visage.

Dans ce qui suit nous allons détailler chaque étape du système de reconnaissance faciale, et nous présenterons les principales difficultés rencontrées.

1.7.1 Détection de visage

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. Nous détaillerons quelques unes de ces méthodes dans la section (2.2). Cependant, les solutions proposées jusqu'à maintenant 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 :

- **La pose** : où les images d'un visage changent en fonction de l'orientation de ce dernier (frontal, 45 degrés, profil).

- **La présence ou absence des composantes structurales** : les caractéristiques faciales tels que la barbe, la moustache, et les lunettes causent une grande variabilité des composantes structurales du visage, notamment au niveau de la forme, de la couleur, et de la taille.
- **Les occultations** : les visages peuvent être partiellement occultés par d'autres objets. En effet, dans une image contenant un groupe de personnes par exemple, des visages peuvent partiellement masquer d'autres visages.
- **Les conditions d'illumination** : des facteurs tels que l'éclairage (distribution de la source de lumière, son intensité, son spectre) et les caractéristiques de l'appareil photographique affectent l'aspect d'un visage dans l'image acquise.

1.7.2 Extraction de caractéristiques du visage

L'extraction des caractéristiques telles que les yeux, le nez, la bouche est une étape prétraitement nécessaire à la reconnaissance faciale. On peut distinguer deux pratiques différentes : la première repose sur l'extraction de régions entières du visage, elle est souvent implémentée avec une approche globale de reconnaissance de visage. La deuxième pratique extrait des points particuliers des différentes régions caractéristiques du visage, tels que les coins des yeux, de la bouche et du nez. Elle est utilisée avec une méthode locale de reconnaissance et aussi pour l'estimation de la pose du visage.

Par ailleurs, plusieurs études ont été menées afin de déterminer les caractéristiques qui semblent pertinentes pour la perception, la mémorisation et la reconnaissance d'un visage humain. Dans [Bru88; She81] par exemple, les caractéristiques pertinentes rapportées sont : les cheveux, le contour du visage, les yeux et la bouche. Cette étude a également démontré le rôle important que joue le nez dans la reconnaissance faciale à partir des images de profil. En effet, dans ce cas de figure, il est évident que la forme distinctive du nez est plus intéressante que les yeux ou la bouche [Bru88]. Dans [She81], les auteurs ont particulièrement établi que la partie supérieure du visage est plus utile pour la reconnaissance faciale que la partie inférieure [She81].

1.7.3 La reconnaissance de visage

Le module de reconnaissance exploite les caractéristiques du visage ainsi extraites pour créer une signature numérique qu'il stocke dans une base de données. Ainsi, à chaque visage de la base est associée une signature unique qui caractérise la personne correspondante. La reconnaissance d'un visage requête est obtenue par l'extraction de la signature requête correspondante et sa mise en correspondance avec la signature la plus proche dans la base de données. La reconnaissance dépend du mode de comparaison utilisé : vérification ou identification (voir section 1.2).

1.8 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.

1.8.1 Changement d'illumination

L'apparence d'un visage dans une image varie énormément en fonction de l'illumination de la scène lors de la prise de vue (voir figure 1.7). Les variations d'éclairage rendent la tâche de reconnaissance de visage très difficile. En effet, le changement d'apparence d'un visage dû à l'illumination, se révèle parfois plus critique que la différence physique entre les individus, et peut entraîner une mauvaise classification des images d'entrée. Ceci a été expérimentalement observé dans Adini et al [Adi97] où les auteurs ont utilisé une base de données de 25 individus. L'identification de visage dans un environnement non contrôlé reste donc un domaine de recherche ouvert. Les évaluations FRVT [Bla01] ont révélé que le problème de variation d'illumination constitue un défi majeur pour la reconnaissance faciale.



Figure 1. 8. Exemple de variation d'éclairage.

1.8.2 Variation de 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 [Bla01; Phi03]. 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 1.8).



Figure 1. 9. Exemples de variation de poses.

1.8.3 Expressions faciales

Un autre facteur qui affecte l'apparence du visage est l'expression faciale (voir figure 1.9). La déformation du visage qui est due aux expressions faciales est localisée principalement sur la partie inférieure du visage. L'information faciale se situant dans la partie supérieure du visage reste quasi invariable. Elle est généralement suffisante pour effectuer une identification. Toutefois, étant donné que l'expression faciale modifie l'aspect du visage, elle entraîne forcément une diminution du taux de reconnaissance. L'identification de visage avec expression faciale est un problème difficile qui est toujours d'actualité et qui reste non résolu. L'information temporelle fournit une connaissance additionnelle significative qui peut être utilisée pour résoudre ce problème [Gro01].



Figure 1. 10. Exemples de variation d'expressions.

1.8.4 Présence ou absence des composants structurels

La présence des composants structurels telle que la barbe, la moustache, ou bien les lunettes peut modifier énormément les caractéristiques faciales telles que la forme, la couleur, ou la taille du visage. De plus, ces composants peuvent cacher les caractéristiques faciales de base causant ainsi une défaillance du système de reconnaissance. Par exemple, des lunettes opaques ne permettent pas de bien distinguer la forme et la couleur des yeux, et une moustache ou une barbe modifie la forme du visage.

1.8.5 Occultations partielles

Le visage peut être partiellement masqué par des objets dans la scène, ou par le port d'accessoire tels que lunettes, écharpe... 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. Gross et al [Gro01] ont étudié l'impact du port de lunettes de soleil, et du cache-nez occultant la partie inférieure du visage sur la reconnaissance faciale. Ils ont utilisé la base de données AR [Mar98]. Leurs résultats expérimentaux semblent indiquer que, dans ces conditions, les performances des algorithmes de reconnaissance restent faibles.

1.9 Conclusion

Dans ce chapitre, nous avons présenté les technologies utilisées dans les systèmes biométriques pour l'identification de personnes. Nous avons aussi donné un aperçu sur les techniques de mesure de leurs performances. Cette étude nous a permis de constater que la reconnaissance de visage suscite de plus en plus l'intérêt de la communauté scientifique, car elle présente plusieurs challenges et verrous technologiques. Enfin, nous avons mis en évidence les différentes difficultés inhérentes à la reconnaissance automatique de visages, ce qui nous a permis de bien définir les problématiques traitées dans cette thèse, notamment l'invariance à l'illumination. Les techniques utilisées aux différentes étapes de la reconnaissance de visage sont détaillées dans le chapitre suivant.

Chapitre 2

Techniques de détection et de reconnaissance de visages

2.1 Introduction

Comme nous l'avons évoqué au chapitre précédent, un système automatique de reconnaissance de visages se décompose en trois sous-systèmes : détection de visage, extraction des caractéristiques et reconnaissance de visages. La mise en oeuvre d'un système automatique et fiable de reconnaissance faciale est un verrou technologique qui n'est toujours pas résolu. Dans ce chapitre, nous présenterons un état de l'art sur les techniques de détection de visage. Nous détaillerons en particulier les méthodes basées sur la couleur de la peau. Nous exposerons aussi les méthodes de reconnaissance de visage 2D et 3D. Enfin, nous terminerons le chapitre par une synthèse des méthodes et techniques étudiées.

2.2 Détection de visages

La détection de visages est la première étape dans le processus de reconnaissance faciale. Son efficacité a une influence directe sur les performances du système de reconnaissance de visages. Il existe plusieurs méthodes pour la détection de visages, certaines utilisent la couleur de la peau, la forme de la tête, l'apparence faciale, alors que d'autres combinent plusieurs de ces caractéristiques.

Les méthodes de détection de visages peuvent être subdivisées en quatre catégories [Yan02] :

2.2.1 Approches basées sur les connaissances acquises

Cette méthodologie s'intéresse aux parties caractéristiques du visage comme le nez, la bouche et les yeux. Elle est basée sur la définition de règles strictes à partir des rapports entre les caractéristiques faciales. Ces méthodes sont conçues principalement pour la localisation de visage. Dans [Kot97], Kotropoulous et Pitas utilisent une méthode à base de règles. Tout d'abord, les caractéristiques du visage sont localisées à l'aide de la méthode de projection proposée par Kanade[Kan73] pour détecter les contours d'un visage. Soit $I(x,y)$ l'intensité de la luminance du pixel (x,y) de l'image $m \times n$, les projections horizontale et verticale de cette image sont définies par $HI(x) = \sum_{y=1}^n I(x,y)$ et $VI(x) = \sum_{x=1}^m I(x,y)$. Le profil horizontal de l'image originale est calculé en premier. Les deux minima locaux sont déterminés, ils correspondent aux bords gauche et droit du visage. Ensuite, le profil vertical est à son tour calculé. Les minima locaux de ce profil vertical correspondent aux positions de la bouche, du nez et des yeux. L'inconvénient de cette méthode est qu'elle n'arrive pas à détecter le visage lorsque ce dernier se trouve sur un arrière plan complexe.

Yang and Huang [Yan94] quant à eux, ont étudié les évolutions des caractéristiques du visage en fonction de la résolution. Quand la résolution de l'image d'un visage est réduite progressivement, par sous-échantillonnage ou par moyenne, les traits macroscopiques du visage disparaissent. Ainsi, pour une résolution faible, la région du visage devient uniforme. Yang et Huang se sont basés sur cette observation pour proposer une méthode hiérarchique de détection de visages. En commençant par les images à faible résolution, un ensemble de candidats de visage est déterminé à l'aide d'un ensemble de règles permettant de rechercher les régions uniformes dans une image. Les candidats de visage sont ensuite vérifiés en cherchant l'existence de traits faciaux proéminents grâce au calcul des minima locaux à des résolutions supérieures. Une caractéristique intéressante de cette technique « descendante » de recherche de zone d'intérêt (informations globales vers des informations plus détaillées) est de réduire le temps de calcul nécessaire par l'utilisation d'images sous-échantillonnées. Malheureusement, cette technique occasionne de nombreuses fausses détections et un taux faible de détection.

2.2.2 Approches basées sur le « Template-matching »

Les templates peuvent être définis soit "manuellement", soit paramétrés à l'aide de fonctions. L'idée est de calculer la corrélation entre l'image candidate et le template. Ces méthodes rencontrent encore quelques problèmes de robustesse liés aux variations de lumière, d'échelle, etc. Sinha [Sin94] [Sin95] utilise un ensemble d'invariants décrivant le modèle du visage. Afin de déterminer les invariants aux changements de luminosité permettant de caractériser les différentes parties du visage (telles que les yeux, les joues, et le front); cet algorithme calcule ainsi les rapports de luminance entre les régions du visage et retient les directions de ces rapports (par exemple, la région 1 est elle plus claire ou plus sombre que la région 2). La figure (6) montre un modèle prédéfini correspondant à 23 relations. Ces relations prédéfinies sont classifiées en 11 relations essentielles (flèches) et 12 relations confirmations (gris). Chaque flèche représente une relation entre deux régions. Une relation est vérifiée si le rapport entre les deux régions qui lui correspond dépasse un seuil. Le visage est localisé si le nombre de relations essentielles et de confirmation dépasse lui aussi un seuil.

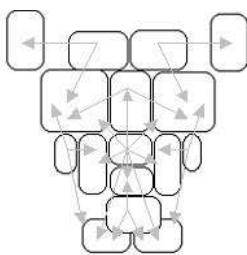


Figure 2. 1 Modèle de visage composé de 16 régions (les rectangles) associées à 23 relations (flèches). [Sin94]

Yuille et al. [Yui92] ont utilisé un template déformable pour modéliser les caractéristiques faciales. Ce template adapte un modèle élastique, connu a priori, aux caractéristiques faciales (e.g., yeux). Dans cette approche les caractéristiques faciales sont décrites par des templates paramétriques. Une fonction d'énergie est définie pour lier les contours, les sommets et les vallées dans l'image d'entrée aux paramètres correspondants dans le template. Le meilleur ajustement du modèle élastique est trouvé en minimisant une fonction d'énergie des paramètres. Bien que leurs résultats expérimentaux démontrent une bonne performance pour le suivi de caractéristiques non rigides, un inconvénient de cette

approche est que le template déformable doit être initialisé dans la proximité de l'objet d'intérêt.

Pour détecter les caractéristiques faciales pour la reconnaissance de visage, Brunelli et Poggio[Bru93] ont utilisé, pour chaque région extraite, un détecteur approprié. Ils se sont aussi inspirés de la méthode de Kanade [Kan73] décrite dans le paragraphe 2.2.1.

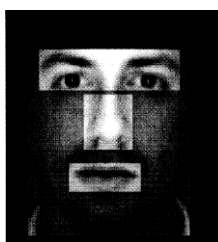


Figure 2. 2. Différentes régions utilisées pour la phase de template matching

Pour les régions yeux, nez et bouche, ils utilisent la direction du gradient vertical et horizontal. La bouche et le nez sont localisés en utilisant des stratégies similaires. La position verticale est déterminée grâce aux standards anthropométriques. D'abord, une estimation fine de leur position réelle est obtenue en cherchant les pics de la projection horizontale du gradient vertical pour le nez, et les vallées de la projection horizontale de l'intensité pour la bouche

La position des sourcils et leur épaisseur peuvent être trouvées par une analyse similaire. La recherche est une fois encore limitée à la fenêtre d'intérêt, juste au-dessus des yeux, et les sourcils sont trouvés en utilisant la carte du gradient vertical. Le détecteur du sourcil cherche les paires de pics du gradient ayant des directions opposées.

2.2.3 Approches basées sur l'apparence

Ces approches appliquent généralement des techniques d'apprentissage automatique. Ainsi, les modèles sont appris à partir d'un ensemble d'images représentatives de la variabilité de l'aspect facial. Ces modèles sont alors employées pour la détection. L'idée principale de ces méthodes est de considérer que le problème de la détection de visage est un problème de classification (visage, non-visage). Une des approches les plus connues de détection de visage est l'*Eigenface* [Tur91], elle consiste à projeter l'image dans un espace et à calculer la distance euclidienne entre l'image et sa projection. En effet, en codant l'image dans un espace, on dégrade l'information contenue dans l'image, puis on calcule la perte

d'information entre l'image et sa projection. Si cette perte d'information est grande (évaluée à partir de la distance, que l'on compare à un seuil fixé a priori), l'image n'est pas correctement représenté dans l'espace : elle ne contient pas de visage. Cette méthode donne des résultats assez encourageants, mais le temps de calcul est très important.

Dans Rowley et al. [Row98], les auteurs proposent un système de détection de visage basé sur la classification par des réseaux de neurones. Leur technique est divisée en deux étapes: la localisation des visages en utilisant un réseau de neurones et la vérification des résultats obtenus. Les auteurs ont construit un réseau de neurones qui, à partir d'une image prétraitée de 20x20 pixels, indique s'il s'agit d'un visage ou non. Le prétraitement consiste à égaliser l'histogramme. L'image est balayée avec des fenêtres de 20x20. Pour détecter les visages de différentes tailles, une analyse multi-résolutions est effectuée. L'extension a aussi été proposée afin de déterminer un réseau de neurones indiquant le degré de rotation d'un visage. Ainsi, le système est capable de détecter des visages ayant subi des rotations dans le plan et de différentes échelles. L'un des avantages des réseaux de neurones est leur robustesse au bruit. Malheureusement, les réseaux de neurones, sont souvent difficiles à construire. Leur structure (nombre de couches cachées pour les perceptrons par exemple) influe beaucoup sur les résultats et il n'existe pas de méthode pour déterminer automatiquement cette structure. La phase d'apprentissage est difficile à mener puisque les exemples doivent être correctement choisis (en nombre et en configuration).

La méthode de Féraud et al [Fer01] est basée sur un type particulier de réseau de neurones qualifié de génératif contraint et entraîné sur les valeurs des images globales. Le réseau est un Perceptron Multicouches auto-associatif et entièrement connecté. Il est conçu pour permettre de mettre en œuvre l'Analyse Composantes Principales (ACP) non-linéaire. Cette approche fournit des taux intéressants de détection de visage (voir le tableau 2.1).

Une autre approche performante qui repose sur les réseaux de neurones est celle de Garcia et al [Gar04] appelée « Convolutional Face Finder (CFF) », elle est basée sur l'utilisation des réseaux de neurones convolutionnels multicouches. Le réseau, entraîné de manière supervisée depuis une base d'apprentissage de visage et non visage, est capable de dériver automatiquement des extracteurs de caractéristiques (produit de convolution) spécialisés. Celui-ci fournit sur la base CMU un taux de détection égal à 90,3%.

Schneiderman et Kanade [Sch00] utilise une information multi-résolution pour les différents niveaux de la transformée en ondelette. La classification en visage non linéaire ou bien en non visage est réalisée en utilisant les statistiques des produits d'histogrammes. Ces derniers sont calculés à partir des exemples de visage et de non visage en utilisant la méthode « AdaBoost learning » [Sch00a] [Li04]. Adaboost est employé pour résoudre trois problèmes fondamentaux : 1) apprentissage effectif des caractéristiques à partir un grand ensemble de caractéristiques. 2) construction de classifieurs faibles, dont chacun est basé sur une des caractéristiques choisies; et 3) le renforcement ou « boosting » des classifieurs faibles pour construire un classifieur fort. Le coût de calcul de cette technique est très élevé.

Une méthode bien connue de détection d'objets complexes tels que les visages est l'utilisation de « classifieurs de Haar » montés en cascade (boostés) au moyen d'un algorithme AdaBoost. Cette méthode est implémentée nativement dans la bibliothèque OpenCV [Bra05] et a été présenté initialement dans Viola et Jones [Vio01]. Le principe de cette méthode est obtenir un algorithme complexe de classification, composé de classifieurs élémentaires qui éliminent au fur et à mesure les zones de l'image qui ne sont pas compatibles avec l'objet recherché. Ces classifieurs binaires reposent sur des primitives visuelles qui dérivent des fonctions de Haar (Haar- like features).

2.2.4 Approches basées sur des caractéristiques invariantes

Ces approches sont utilisées principalement pour la localisation de visage. Les algorithmes développés visent à trouver les caractéristiques structurales existantes même si la pose, le point de vue, ou la condition d'éclairage changent. Puis ils emploient ces caractéristiques invariables pour localiser les visages. Nous pouvons citer deux familles de méthodes appartenant à cette approche : Les méthodes basées sur la couleur de la peau qu'on détaillera dans la section suivante, et les méthodes basées sur les caractéristiques de visage; elles consistent à localiser les cinq caractéristiques (deux yeux, deux narines, et la jonction nez/lèvre) pour décrire un visage typique.

L'algorithme développé par De Silva et al. [DeS95] est un exemple typique des méthodes basées sur les caractéristiques. Leur algorithme utilise tout d'abord une hypothèse sur la position du haut du visage ensuite l'algorithme de recherche parcourt le visage de haut en bas afin de trouver l'axe des yeux « eye-plane » caractérisé par une augmentation soudaine

de la densité de contours (mesurée par le rapport noir/blanc le long des plans horizontaux). La longueur entre le haut du visage et le plan de l'oeil est alors utilisée comme une longueur de référence pour construire un « template » facial flexible. Ce « template » couvrant des caractéristiques telles que les yeux et la bouche est initialisé à partir de l'image d'entrée. La forme initiale du « template » est obtenue en utilisant la longueur anthropométrique en respectant la longueur de référence. Le template flexible est alors ajusté par rapport aux positions finales des caractéristiques en utilisant un algorithme de réglage fin qui emploie une fonction de coût basée contour. Les auteurs rapportent un taux de 82% de réussite pour la détection de toutes les caractéristiques faciales dans le cas de visages quasi-frontaux ($<30^\circ$) sur un fond homogène. Bien que l'algorithme réussisse à détecter les caractéristiques d'ethnies différentes puisqu'il ne se base pas sur les informations de niveaux de gris et de couleur, il n'arrive pas cependant à détecter correctement ces caractéristiques si l'image du visage contient des lunettes ou bien si les cheveux couvrent le front.

Par ailleurs, Yow et Cipolla [Yow97] ont développé une méthode basée sur l'extraction de points d'intérêts (les maxima locaux de la dérivée seconde d'un filtre gaussien). Ces points sont ensuite utilisés pour réaliser une détection de contours qui sont à leur tour groupés en régions. Le regroupement des contours est basé sur un critère de proximité et un critère de similarité en orientation et en épaisseur. A partir de chaque région, l'algorithme définit alors un vecteur pour lequel il calcule la moyenne et la matrice de covariance par rapport aux différents vecteurs des modèles. Le critère d'appartenance à un élément du visage s'appuie sur la distance de Mahalanobis: Les différents candidats sont alors groupés en se basant sur un modèle de connaissance indiquant leur position relative. Chaque composant du visage est enfin analysé avec un réseau bayésien. L'intérêt de cette méthode est qu'elle peut détecter des visages dans diverses poses. On a un taux de réussite de 85%.

Détecteur de visages	CMU	CMU-125	MIT	MIT-20
Rowley et al. [Row98],	86,2%/23		84,5%/8	
Schneiderman et Kanade [Sch00]		94,4%/65		
Féraud et al [Fer01]	86%/8			
Viola et Jones [Vio01]	88,4%/31		77,8%/5	
Garcia et al [Gar04]	90,3/8	90,5/8	90,1/7	90,2/5

Tableau 2. 1. Extrait de [Gar04]. Evaluation des performances des principales méthodes de détection de visage (Taux de détection / Nombre de Faux Positifs) sur les bases CMU et MIT.

Dans le cadre de cette thèse, nous nous intéressons en particulier aux techniques de détection de visages basées sur l'analyse de la couleur de la peau. La section suivante, présentera une étude détaillée sur ces techniques, ce qui nous permettra de situer notre approche par rapport aux travaux existants.

2.2.5 Détection de visages basée sur l'analyse de la couleur de la peau

Les méthodes de détection basées sur l'analyse de la couleur de la peau sont des méthodes efficaces et rapides. Elles réduisent l'espace de recherche de la région visage dans l'image. De plus, la couleur de la peau est une information robuste face aux rotations, aux changements d'échelle, et aux occultations partielles. Plusieurs espaces couleur peuvent être utilisés pour détecter, dans l'image, les pixels qui ont la couleur de la peau. L'efficacité de la détection dépend essentiellement de l'espace couleur choisi. Les espaces couleur les plus utilisés sont :

- *L'espace RVB*, mis au point en 1931 par la Commission Internationale de l'Eclairage (CIE). Il consiste à représenter l'espace des couleurs à partir de trois rayonnements monochromatiques de couleurs : Rouge-Vert-Bleu. Cet espace correspond à la façon dont les couleurs sont généralement codées informatiquement, ou plus exactement à la manière dont les écrans à tubes cathodiques (ordinateurs, TV) représentent ces couleurs.
- *L'espace HSL* (Hue, Saturation, Luminance), appelé aussi TSL (Teinte, Saturation, Luminance) en Français, s'appuie sur les travaux du peintre Albert H. Munsell. C'est un modèle de représentation dit "naturel", car il est proche de la perception physiologique de la couleur par l'œil humain. En effet, le modèle RGB aussi adapté

soit-il pour la représentation informatique de la couleur ou bien l'affichage sur les périphériques de sortie, ne permet pas de sélectionner facilement une couleur. Le modèle HSL consiste à décomposer la couleur selon des critères physiologiques :

- la teinte (en Anglais Hue), correspondant à la perception de la couleur,
- la saturation, décrivant la pureté de la couleur, c'est-à-dire son caractère vif ou terne,
- la luminance, indiquant la quantité de lumière de la couleur, c'est-à-dire son aspect clair ou sombre.

Il existe d'autres modèles naturels de représentation proches du modèle HSL :

- HSB : Hue, Saturation, Brightness soit en Français Teinte, Saturation, Brilliance. La brillance décrit la perception de la lumière émise par une surface.
- HSV : Hue, Saturation, Value soit en Français Teinte, Saturation, Valeur.
- HSI : Hue, Saturation, Intensity soit en Français Teinte, Saturation, Intensité.
- HCI : Hue, Chrominance, Intensity soit Teinte, Chrominance, Intensité.
- *Le modèle YCrCb* est un signal non-linéaire codé à partir du signal RVB. Le paramètre Y représente la luminance (c'est-à-dire l'information en noir et blanc), tandis que Cr et Cb permettent de représenter la chrominance, c'est-à-dire l'information couleur.
- *Le modèle colorimétrique $L^*a^*b^*$* (aussi connu sous le nom de CIELab), dans lequel une couleur est repérée par trois valeurs :
 - L^* , la luminance, exprimée en pourcentage (0 pour le noir à 100 pour le blanc).
 - a^* et b^* deux gammes de couleur allant, respectivement, du vert au rouge et du bleu au jaune avec des valeurs allant de -120 à +120.

Le modèle $L^*a^*b^*$ couvre ainsi l'intégralité du spectre visible par l'œil humain et le représente de manière uniforme. Il permet donc de décrire l'ensemble des couleurs visibles indépendamment de toute technologie graphique.

Les techniques de détection du visage basées sur la couleur de la peau peuvent être classifiées en quatre catégories que nous allons maintenant aborder: les méthodes

explicités, les méthodes non paramétriques, les méthodes paramétriques, et les méthodes semi paramétriques. Toutes ces approches pratiquent une phase d'apprentissage sur un nombre d'images représentatives pour calculer une densité de probabilité de la couleur peau.

2.2.5.1 Les méthodes explicites

Les méthodes explicites utilisent des règles de décision empiriques et/ou statistiques [Vez03] pour la détection des pixels ayant la couleur de la peau. Les images doivent être acquises dans un environnement contrôlé (i.e. un éclairage réglable) [Yan98a]. Une méthode explicite est une méthode de classification qui consiste à définir explicitement les frontières de la région peau (cluster) dans l'espace couleur utilisé.

Dans [Pee03], par exemple, la peau est classifiée dans l'espace couleur *RGB* en utilisant les règles suivantes :

$$\begin{aligned} R > 95 \text{ et } G > 40 \text{ et } B > 20 \text{ et} \\ \max\{R, G, B\} - \min\{R, G, B\} > 15 \text{ et} \\ |R - G| > 15 \text{ et } R > G \text{ et } R > B \end{aligned} \quad (2.1)$$

Par ailleurs, Chai et Ngan [Cha99] ont proposé un algorithme de segmentation de visage dans lequel ils ont employé les deux plans *CbCr* du modèle couleur *YCrCb* pour déterminer les régions ayant la couleur peau. Ils ont utilisé la base de données de visages ECU [ECU]. Ils ont trouvé que les valeurs de pixels dans les domaines $D_{Cb} = [77, 127]$ et $D_{Cr} = [133, 173]$ définissent bien les pixels peau. De même, Garcia et Tziritas [Gar99] ont segmenté la peau en utilisant, huit plans dans l'espace *YCbCr*, ou six plans dans l'espace *HSV*.

L'avantage de ces méthodes réside dans la simplicité des règles de détection de la peau qu'elles utilisent, ce qui permet une classification rapide. Cependant, leur problème principal est la difficulté de déterminer empiriquement un espace couleur approprié ainsi que des règles de décision adéquates qui assurent un taux de reconnaissance élevé.

Une méthode utilisant des algorithmes d'apprentissage a été proposée pour résoudre ces problèmes [Gom02]. Les auteurs commencent par choisir un espace *RGB* normalisé (où $r = R / R+B+G$, $g = G / R+B+G$ et $b = B / R+B+G$ et la somme des trois composantes normalisées $r+g+b=1$) sur lequel ils appliquent un algorithme d'induction constructive

afin de créer de nouveaux ensembles d'attributs pour les composantes RGB. Une règle de décision, semblable à l'équation (2.1) qui réalise la meilleure identification possible, est estimée pour chaque ensemble d'attributs. Ils ont obtenu des résultats meilleurs que ceux qui sont obtenus avec un classifieur de Bayes défini dans l'espace RGB.

2.2.5.2 Les approches non paramétriques

Les approches non paramétriques ne dépendent pas de la forme de la fonction de distribution de la teinte. Elles utilisent les histogrammes couleur 2D et 3D pour représenter la distribution (ou densité de probabilité) de la tonalité de la peau dans un espace couleur. L'avantage d'utiliser les histogrammes couleur est qu'ils sont robustes (invariants) aux occultations et aux changements de point de vue. Ils peuvent ainsi différencier un grand nombre d'objets [Yan98a] dans une tâche de reconnaissance.

En général, les approches non paramétriques se déroulent en trois étapes :

- Construire les histogrammes de couleur de peau et de non peau à partir de l'espace couleur.
- Calculer la probabilité conditionnelle pour chaque couleur de peau et de non peau.
- Utiliser la règle de Bayes afin de calculer la probabilité pour qu'un pixel corresponde à la classe peau. Ce processus, permet ainsi de créer une carte de probabilité de la couleur peau.
- Un seuil de classification est enfin déterminé à partir de la relation entre les détections correctes et les détections fausses données par la courbe ROC (Receiver Operating Characteristic). Il est appliqué à la carte de probabilité et permet d'extraire les régions peau.

2.2.5.3 Les approches paramétriques

Les approches non-paramétriques basées sur les histogrammes couleur exigent un grand espace mémoire. De plus, leurs performances dépendent directement des images d'apprentissage. Le besoin d'un modèle plus compact pour représenter la peau, et qui peut être généralisé avec moins de données d'apprentissage, a poussé les chercheurs à développer des modèles paramétriques de distribution de peau. Cette dernière est souvent

représentée sous la forme d'un modèle gaussien simple (MGS) parfois elliptique, ou par un mélange de modèles gaussiens (GMM) pour traduire son caractère multimodal.

Plusieurs travaux sur la modélisation de la distribution de la couleur de peau ont utilisé un mélange de Gaussiennes défini par :

$$P(C; \mu, \Sigma) = \sum_{k=1}^n \alpha_k \frac{1}{\sqrt{(2 * \pi)^d |\Sigma_k|}} \exp^{-0.5(C - \mu_k)^T (\Sigma_k)^{-1} (C - \mu_k)} \quad (2.2)$$

Où C est la représentation des espaces couleur utilisés, n le nombre de gaussiennes utilisées, α le poids de la k^{ieme} gaussienne, Σ la matrice de covariance, μ le vecteur moyen, et d la dimension des données de C .

Pour une gaussienne simple (uni-modale), les valeurs de α et de n sont égales à 1. Dans ce cas, les paramètres peuvent être estimés en utilisant le maximum de vraisemblance, alors que pour le mélange de gaussiennes ils sont déterminés en utilisant l'algorithme *EM* (Expectation - Maximization algorithm) [Dem77][Jon02][Yan98b].

Le choix du nombre n de gaussiennes dépend énormément des données d'apprentissage et du choix de l'espace couleur utilisé. Plusieurs techniques ont utilisé des valeurs de n comprises entre 2 et 16.

Yang and Ahuja [Yan99] ont utilisé deux gaussiennes dans l'espace couleur LUV. Leurs tests ont démontré que le modèle MGS n'étaient pas suffisants pour modéliser la distribution de la peau avec la base de données de Michigan. Greenspan et al. [Gre01] ont démontré qu'un modèle MMG est une représentation robuste qui s'adapte aux différents espaces couleur et aux variations d'illumination. Leur modèle MMG contient deux composantes, la première modélise la distribution de la lumière ordinaire de la couleur de peau, alors que la deuxième composante modélise la distribution des régions les plus lumineuses de la peau.

En étudiant la distribution de la couleur de peau et de non-peau dans plusieurs espaces couleurs, Lee et Yoo [Lee02] ont conclu que les régions de couleur de peau ont approximativement une forme elliptique qui ne peut pas être modélisée par un MGS. En effet, en raison de l'asymétrie du cluster de peau par rapport au pic de densité, l'utilisation du modèle gaussien symétrique (MGS) mène à un taux élevé de faux positifs. Afin de résoudre ce problème, ils ont proposé donc un nouveau modèle qu'ils ont appelé « modèle

elliptique de frontière ». Les auteurs ont comparé leur modèle avec les modèles MGS et MMG à six composantes, appliqués sur la base de données de Compaq [Jon99]. Ils ont obtenu des performances légèrement meilleures. Cependant, l'inconvénient du modèle elliptique de frontière réside dans le fait que son utilisation est limitée à la classification binaire.

Hsu [Hsu02] propose de combiner une technique de correction de couleur avec la détection de la couleur de peau, pour localiser le visage dans une image. La correction de couleur permet d'éliminer l'effet de la réflexion et de la variation de l'illumination dans l'image. La technique de compensation d'éclairage utilise ce qu'on appelle un « blanc de référence ». Elle est basée sur l'hypothèse qu'une image contient toujours du blanc. Les pixels blancs sont reconnus en utilisant une correction non-linéaire Gamma de la valeur de luminance. Si le pourcentage des pixels blancs dépasse 5 % sur 100 pixels, on applique alors une correction de couleur sur l'image. Cette correction s'applique sur les coefficients des trois axes RGB.

Enfin, pour détecter la couleur de peau on utilise un MGS dans l'espace modifié CbCr, qui est obtenu à partir d'une transformation non linéaire appliquée à l'espace YCbCr. L'inconvénient de cette méthode de correction réside dans la difficulté à définir les hypothèses de départ [Mar03].

2.2.5.4 Approches semi-paramétriques

Décrit par Kohonen au début des années 80, le réseau SOM (Self-Organizing Map) [Koh97] est un des plus populaires réseaux de neurones non supervisé. Il est principalement utilisé pour classifier des données de grande dimension, mais il fonctionne tout aussi bien pour des données de faible dimension. L'objectif du réseau SOM est de classifier les pixels d'entrée selon qu'ils correspondent ou non à des pixels de peau, ce qui permet au final d'extraire la région du visage. Brown et al. [Bro01] entraînent deux réseaux SOM pour apprendre, à partir de 500 images, la distribution des pixels de la couleur peau et la couleur non peau. Les performances du SOM ont été testées sur l'ensemble des images d'apprentissage/test de la base de données Compaq [Jon99]. Plusieurs espaces couleur (RGB normalisé, Teinte-Saturation, TLS) ont été utilisés avec le détecteur SOM. Les résultats ont montré que les performances des détecteurs de peau SOM ne dépendent pas des espaces couleurs utilisés, à l'inverse du modèle MMG. Par ailleurs, les

performances du détecteur SOM sont inférieures à celles qui sont obtenues par la méthode basée sur les histogrammes RGB, développée dans [Jon99]. Néanmoins, le détecteur SOM nécessite moins de paramètres et il est plus efficace pour les applications en temps réel.

Certains auteurs ont également proposé un apprentissage automatique des couleurs représentant la peau à l'aide de réseaux neuronaux. Ceux-ci peuvent être entraînés à partir d'échantillons de pixels représentant la peau (et non-peau), préalablement convertis dans l'espace de couleur YCrCb. Par ailleurs, d'autres espaces de couleurs ont aussi été utilisés dans ce contexte. Kakumanu et al. [Kak04][Kak06] ont employé un réseau de neurones pour modéliser la constance de la couleur « the color constancy ». Leur réseau de neurones possède trois couches et permet d'estimer l'illumination de la peau. Il prend en entrée un histogramme à deux composants r et g (voir section 2.2.1.1) et donne en sortie les illuminations de la peau dans l'espace r, g . Ce réseau de neurones est entraîné sur une base de données de 255 images, et testé sur 71. Les images utilisées fournissent un grand choix d'illuminations, de différents arrières plans et de sources de lumière non blanches. Enfin, une technique de seuillage simple est utilisée pour détecter la peau à partir de ce réseau de neurones.

2.2.5.5 Autres approches

Sobottka et Pitas [Sob98] proposent une méthode de localisation et d'extraction des caractéristiques du visage qui utilise à la fois la forme et la couleur [Sob96]. La segmentation de l'image se fait dans l'espace HSV pour localiser la région couleur de la peau. Les composantes connexes sont ensuite déterminées en appliquant un algorithme de croissance de régions avec une résolution grossière. Pour chaque composante connexe, l'algorithme ajuste une ellipse afin de déterminer la région candidate qui correspond au visage. Enfin, une analyse plus fine des caractéristiques à l'intérieur de cette région permet de conclure sur la présence d'un visage ou non.

Dans [Ter98a][Ter98b], un modèle gaussien de la couleur de peau est utilisé pour classifier les pixels de couleurs de peau. Afin de caractériser la forme des « clusters » dans l'image binarisée, un ensemble de 11 moments géométriques d'ordre inférieur est calculé en utilisant la transformée de Fourier et la transformée radiale de Mellin. Afin de détecter la région visage, un réseau de neurones est entraîné à l'aide des moments géométriques

extraits. L'expérimentation de cette méthode a démontré un taux de détection de 85% sur une base de test de 100 images.

Belaroussi et al. [Bel05] proposent de combiner trois détecteurs pour localiser le visage :

- un détecteur anthropomorphique basé sur un modèle d'apparence neuronal. Il s'agit d'un réseau de neurones auto-associateur (réseau Diabolo déjà utilisé en reconnaissance d'écriture). Ce réseau est entraîné à reconstruire la classe "visage" (pendant l'apprentissage, l'entrée sert de sortie désirée) en réalisant une compression des données (analyse en composantes principales non linéaire) dans sa couche cachée comportant un faible nombre de cellules. Une image de non-visage sera en principe mal compressée et donnera une erreur de reconstruction plus importante.
- un détecteur géométrique basé sur la Transformation de Hough Généralisée (THG). L'ellipse est une forme géométrique simple qui permet de modéliser grossièrement un visage. Une THG est donc réalisée sur l'image des orientations de gradients afin de détecter une ellipse verticale d'excentricité donnée. Il en résulte un tableau de vote dont le maximum correspond à la position dans l'image du point le plus susceptible d'être le centre de l'ellipse.
- un détecteur colorimétrique basé sur une modélisation statistique de la teinte chair par mélange de gaussiennes (MMG) dans l'espace YCbCr.

Le traitement parallèle d'une image par les trois détecteurs produit trois cartes de probabilité qui sont combinées linéairement. Le maximum (cas mono-visage) ou les n premiers maxima (cas multi-visages) de la combinaison donne la localisation. Les expériences, conduites sur la base ECU (3000 images), montrent que la combinaison des détecteurs permet d'atteindre des taux de localisation supérieurs à 85%.

2.2.6 Comparaison des différentes approches

Les performances des différentes méthodes de modélisation de la couleur de la peau ont été évaluées sous des conditions identiques. Malheureusement, la plupart des méthodes de détection de la peau fournissent des résultats obtenus sur leurs propres bases de données qui ne sont pas disponibles. La base d'apprentissage et de test la plus utilisée pour la détection de la peau est la base de données Compaq [Jon99].

Dans le tableau (2.1) les performances des différentes méthodes appliquées à cette base de données, sont présentés. Il donne le taux des vrais positifs (TP) et le taux des faux positifs (FP). Bien que les méthodes utilisent la base de données de manière différente pour définir les images d'apprentissage et les images de test, et emploient différentes stratégies d'apprentissage, ce tableau donne une image assez fidèle des performances obtenues par ces méthodes [Kak06].

Méthodes	Espace couleur	TP	FP
Bayes SPM [Jon99]	RGB	90%	14.2%
Bayes SPM [Bra00]	RGB	93.4%	19.8%
MMG [Jon99]	RGB	90%	~15.5%
SOM [Bro01]	TLS	78%	32%
Elliptical boundary model [Lee02]	CIE-xy	90%	20.9%
MGS [Lee02]	CbCr	90%	33.3%
MMG [Lee02]	IQ	90%	30%
Thresholding of I axis[Bra00]	YIQ	94.7%	30.2%
MGS [Hsu02]	YCbCr	96%	

Tableau 2. 2 Les performances des différentes méthodes de détection de peau.

2.3 Techniques 2D de reconnaissance de visage

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 reconnaissance faciale est de rivaliser, voir même dépasser, les capacités humaines de reconnaissance. Les résultats fondamentaux de ces études ont été synthétisés dans un papier publié en 2006 par [Sin06], ils se résument comme suit :

- Les humains peuvent reconnaître des visages familiers dans des images de faible résolution.
- La capacité de tolérer les dégradations des images augmente avec la familiarité
- Les informations haute fréquence seules, soit les contours, sont insuffisantes pour obtenir une reconnaissance faciale performante
- Les caractéristiques faciales sont traitées de manière holistique
- Parmi les différentes caractéristiques faciales, les sourcils sont les plus importants pour la reconnaissance
- La forme du visage est généralement codée de manière caricaturale
- La pigmentation du visage est aussi importante que sa forme
- La couleur joue un rôle important spécialement lorsque la forme est dégradée
- Les changements d'illumination influencent la capacité de généralisation
- Le mouvement des visages semble faciliter la reconnaissance de manière conséquente
- Le système visuel progresse d'une stratégie locale vers une stratégie holistique au cours des premières années de la vie
- Identité faciale et expressions sont traitées par des systèmes séparés

C'est ce qui explique l'abondance et la diversité de la littérature dans ce domaine. L'article de Tana et al. [Tan06] 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.

2.3.1 Approches holistiques ou globales

Ces méthodes identifient un visage en utilisant l'image entière de ce dernier comme entrée du système de reconnaissance. Chaque image de 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. L'espace I contenant tous les vecteurs images de visages est appelé espace images. L'avantage de cette représentation est qu'elle préserve implicitement les informations de texture et de forme nécessaire pour la reconnaissance de visages. De plus, elle permet une meilleure capture de l'aspect global du visage que les représentations locales [OTo93]. Toutefois, son inconvénient majeur réside dans la

dimension très grande de l'espace image qu'elle nécessite [Jai82, Jai87, Rau91], ce qui rend très difficile la classification.

Ainsi, une image 100×100 , par exemple, est représentée par un vecteur de dimension 10^4 [Dui95]. Comme le nombre d'images d'apprentissage pour chaque personne doit être au moins égal à dix fois la dimension du vecteur [Jai82], il faut 10^5 images par personne, nombre assez exorbitant. En pratique, on n'a pas besoin de tant de photos pour développer un modèle fidèle pour l'apparence du visage d'une personne. Des techniques de réduction de dimension sont généralement employées. Une des techniques les plus utilisées dans l'identification de visage est l'Analyse en Composantes Principales (ACP).

Une méthode très populaire, basée sur la technique ACP, est la méthode « eigenface » [Sir87, Tur91]. Son principe est le suivant : étant donné un ensemble d'images de visages exemples, il s'agit tout d'abord de trouver les composantes principales de ces visages. Ceci revient à déterminer les vecteurs propres de la matrice de covariance formée par l'ensemble des images exemples. Chaque visage exemple peut alors être décrit par une combinaison linéaire de ces vecteurs propres. Pour construire la matrice de covariance, chaque image de visage est transformée en vecteur. Chaque élément du vecteur correspond à l'intensité lumineuse d'un pixel. Cette méthode sera présentée avec davantage de détails dans le chapitre 5.

Dans [Wan03], les auteurs ont démontré que la matrice de covariance C peut s'écrire :

$$C = C_I + C_E \quad (2.12)$$

C'est-à-dire qu'elle est égale à la somme de la matrice de dispersion intra-personne C_I et la matrice de dispersion inter-personne C_E . Dans le cas d'un seul exemple d'apprentissage par personne, $C_I = 0$, et donc l'équation (2.12) se réduit à C_E .

L'eigenspace estimé à partir de la matrice C_E seulement n'est pas fiable, parce qu'il ne peut pas différencier, de manière efficace, l'erreur d'identification des autres erreurs dues à la transformation et au bruit [Wan03]. Pour illustrer l'influence du nombre d'exemples d'apprentissage par personne sur les performances de la reconnaissance, les auteurs ont utilisé la base de données ORL [Sam94] comme base de test. La base de données ORL contient des images de 40 individus, chacun étant enregistré sous 10 vues différentes. Dans leur expérimentation, les auteurs ont fixé le nombre de visages de test. Par contre, ils ont fait varier le nombre de visages d'apprentissage. Ainsi, pour chaque personne, ils ont utilisé

la dernière image (Figure 2.1) pour le test et ont choisi aléatoirement les n premières images ($n \leq 9$) pour l'apprentissage.



Figure 2.3. Les dix vues d'une personne dans la base de données ORL.

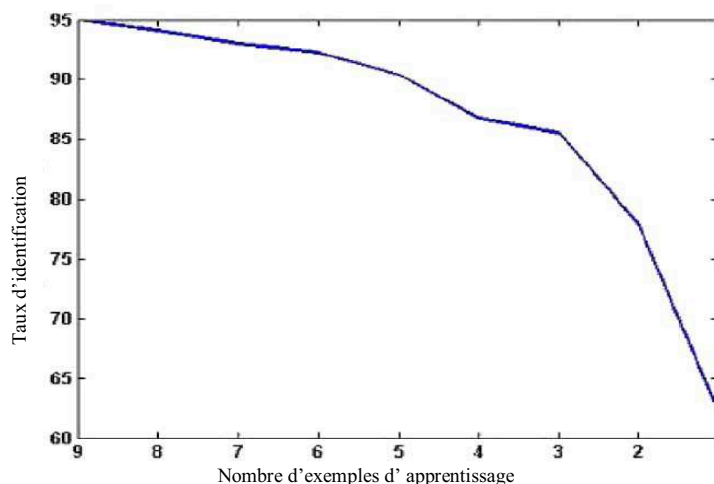


Figure 2.4. Taux d'identification moyen en fonction du nombre d'exemples d'apprentissage par personne [Wan03].

Cette procédure a été répétée vingt fois. La figure 2.2 montre que la performance de la méthode eigenface baisse avec la diminution du nombre d'exemples d'apprentissage pour chaque personne. Dans le cas extrême, si seulement un exemple d'apprentissage par personne est utilisé, le taux d'identification moyen de l'eigenface tombe en dessous de 65 %. Ce taux atteint 95 % quand on utilise neuf exemples d'apprentissage par personne.

Pendant la dernière décennie, plusieurs chercheurs ont développé diverses extensions de la méthode « eigenface », comme l'eigenface probabiliste [Mog97], les méthodes basées sur l'Analyse Discriminante Linéaire (ADL) [Bel97, Zha99, Lu03], la Machine Vecteur Support (SVM) [Phi98], la méthode de la ligne caractéristique [Li99], et Laplacianfaces [He05]. Ces approches possèdent des performances supérieures à celles qui sont données par la méthode eigenface basique. 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 peuvent être résumées comme suit [Tan06] :

- 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. Cependant, si un seul exemple d'apprentissage par personne est utilisé, c'est-à-dire si les variations intra classes nulles, alors les performances de l'ADL deviennent faibles par rapport à celles qui sont données par l'eigenface [Mar01]. Afin de remédier à ce problème, Zhao et al. [Zha99] ont proposé de remplacer la matrice de dispersion intra-personne par une matrice constante. Ainsi, la méthode basée ADL se réduit alors à la méthode eigenface.
- L'approche Probabiliste [Mog97] 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.
- "Evolution pursuit" [Liu00] et « Laplacianfaces » [He05] dépendent largement du nombre d'exemples d'apprentissage par personne. Dans le cas d'un exemple par personne, ces méthodes se ramènent encore à la méthode « eigenface ».
- Les méthodes SVM [Phi98] et ligne de caractéristiques [Li99] sont des méthodes de classification dans l'«eigenspace ». Si uniquement un exemple par personne est utilisé, alors aucune des deux méthodes ne marche.

Comme illustré ci-dessus, on ne peut pas s'attendre à obtenir de bonnes performances en généralisant la technique eigenface lorsque celle-ci utilise un seul exemple par personne. Cependant, cette méthode s'adapte très bien pour la classification de données de grande dimension.

2.3.1.1 Les nouvelles extensions de l'Analyse des Composantes Principales

Wu et Zhou présentent une nouvelle méthode nommée $(PC)^2A$, pour enrichir l'information sur l'espace image [Wu02]. Ils utilisent une méthode de projection pour la détection de visage [Kot97] dont le principe est le suivant :

Soit $I(x,y)$ l'intensité du pixel (x,y) de l'image I de dimension $m \times n$. Les projections horizontale et verticale de cette image sont respectivement définies par :

$$HI(x) = \sum_{y=1}^n I(x,y) \quad (2.13)$$

$$VI(y) = \sum_{x=1}^m I(x,y) \quad (2.14)$$

Les deux projections reflètent la distribution des caractéristiques faciales saillantes utiles pour l'identification de visage. Elles permettent de générer une carte de projection (figure 2.3b). L'image obtenue à partir de ces projections est ensuite combinée avec l'image originale pour générer une image plus riche en information (figure 2.3c). Cette dernière est définie par :

$$M_p(x,y) = HI(x)VI(y)/\bar{J} \quad (2.15)$$

où \bar{J} est l'intensité moyenne de l'image.

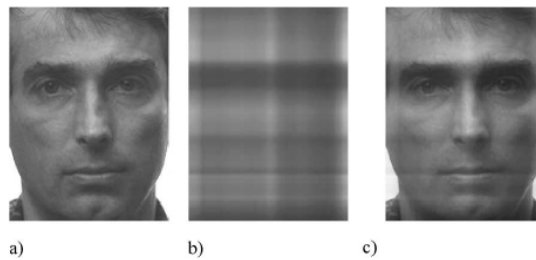


Figure 2. 5. (a) image originale (b) carte de projection ; (c) image combinée.

La technique eigenface est utilisée pour la reconnaissance de visages. Cette méthode a été évaluée sur un sous-ensemble de la base de données FERET [Phi00]. Une seule image de visage a été utilisée lors de la phase d'apprentissage. Les auteurs ont rapporté une précision de 3 à 5 % supérieure à la méthode eigenface standard, et en limitant de 10 à 15 % la taille du vecteur eigenfaces.

Par ailleurs, l'estimation de la matrice de covariance, lorsque la dimension des données est faible, est un problème qui n'est toujours pas résolu par la méthode eigenface standard. Yang et al. [Yan04] ont introduit une nouvelle technique qu'ils ont appelé Analyse en Composantes Principales Bidimensionnelle (ACP2D). Cette méthode applique directement l'ACP sur les images 2D plutôt que sur un seul vecteur. Les auteurs utilisent pour cela une matrice de covariance généralisée, calculée à partir des lignes de l'image originale. Cette

modélisation permet de réduire le coût calculatoire et l'instabilité numérique lors de la construction du modèle.

Les méthodes détaillées ci-dessus traitent le cas d'un seul exemple d'apprentissage. Les problèmes de variations d'expression, d'illumination ou de changement de pose ne sont pas explicitement traités. Jung et al. [Jun04] ont développé un système d'authentification pour traiter ces problèmes. Leur idée est de synthétiser un ensemble d'images corrompues, générées à partir d'une image originale. Ils ont utilisé un modèle de bruit avec trois paramètres contrôlant respectivement, le degré de contraste, la brillance et le flou Gaussien.

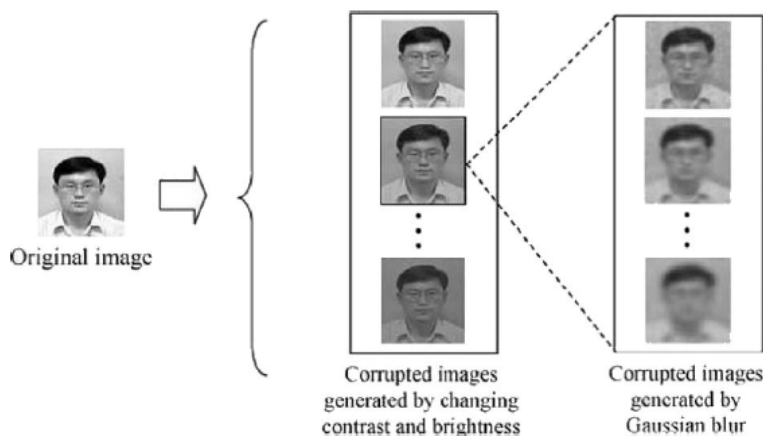


Figure 2. 6. Synthèse de nouvelles images avec différentes dégradations [Jun04]

La figure 2.4 montre un exemple d'images corrompues générées à partir d'une image originale en variant les valeurs des paramètres de bruit. Les auteurs ont utilisé 137 images de visage de cartes d'identité ayant une résolution de 300 dpi pour évaluer leur méthode. Les résultats expérimentaux montrent que le taux d'erreur est seulement de 1.32 %, indiquant que la méthode peut significativement améliorer la similitude entre les images dégradées et les images d'apprentissage.

2.3.1.2 Les nouvelles extensions Fisherface

La méthode des fisherfaces que nous décrivons dans cette section est sans doute la plus connue des approches utilisant l'Analyse Discriminante Linéaire (ADL) dans le contexte de la reconnaissance de visages.

L'approche fisherface [Bel97] est une extension de la méthode eigenface. Son but est de trouver un sous-espace plus discriminant dans l'eigenspace pour la reconnaissance de visages. Cependant, l'approche fisherface ne marche pas lorsqu'un seul exemple est utilisé pour l'apprentissage, car dans ce cas aussi la dispersion intra-classes est nulle.

Afin de résoudre ce problème, Wang et al. [Wan05] ont proposé une méthode qui calcule la dispersion intra-classes à partir de plusieurs sujets différents. Les auteurs justifient leur choix par le fait que l'être humain présente une dispersion intra-classes semblable. Un ensemble générique d'images d'apprentissage, avec des exemples multiples par personne, est collecté. Il est utilisé pour calculer la dispersion intra-personne. Une méthode semblable à celle de Fischer est ensuite implémentée, elle permet de déterminer les vecteurs propres les plus discriminants pour l'identification. Cette méthode a été évaluée sur un sous-ensemble de la base de données FERET avec 256 sujets. Une marge de performance supérieure de 10 % sur la méthode eigenface traditionnelle a été obtenue.

En utilisant l'Analyse en Composantes Principales Bidimensionnelles [Yan04], Visani et Garcia [Vis05] ont proposé une nouvelle méthode basée sur l'algorithme ADL Bilinéaire. Ils ont testé leur technique sur la base Asian Face Image Database PF01, qui contient des prises de vues de 75 personnes dont aucune ne porte de lunettes, avec des conditions d'illumination neutres. Les images sont normalisées. La base d'apprentissage contient quatre vues par personne pour 60 des 75 personnes de la base, avec une expression faciale neutre et une pose frontale. Le taux de reconnaissance rapporté par les auteurs est supérieur à celui de l'eigenface, de fisherface, et de l'ACP2D.

2.3.1.3 Discussion

Bien que les méthodes holistiques aient eu beaucoup de succès, leur inconvénient majeur réside dans le fait qu'elles utilisent uniquement des photos 2D d'apparence faciale. Or, on sait qu'une telle représentation est sensible aux changements d'expression, d'illumination et de poses. Une manière d'éviter ce problème consiste à utiliser des représentations faciales locales. En effet, les caractéristiques locales ne sont généralement pas aussi sensibles aux changements d'apparence que les caractéristiques globales.

2.3.2 Méthodes locales

Les méthodes locales utilisent les caractéristiques faciales locales pour la reconnaissance de visage. Elles sont relativement matures comparées aux méthodes holistiques [Law97, Bru93, Wis97, Man92, Kel70]. Dans ces méthodes, le visage est représenté par un ensemble de vecteurs caractéristiques de dimensions faibles, plutôt que par un seul vecteur de grande dimension.

Nous pouvons classifier les méthodes locales en deux catégories : les méthodes basées sur les caractéristiques locales : extractions et localisation des points caractéristiques, et les méthodes basées sur les apparences locales : partitions des images de visage en région caractéristiques.

2.3.2.1 Méthodes basées sur les caractéristiques locales

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.

2.3.2.1.1 Approches géométriques

Elles sont basées sur l'extraction de la position relative des éléments qui constituent le visage (tel 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). Au début des années 1990, Brunelli et Poggio [Bru93] ont décrit un système de reconnaissance faciale qui extrait automatiquement 35 caractéristiques géométriques du visage. La similitude est calculée à l'aide de classifieurs de Bayes. Un taux d'identification de 90 % sur une base de données de 47 sujets a été rapporté par les auteurs. 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, etc.
- 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.

2.3.2.1.2 Approches basées sur les graphes

Plutôt que d'utiliser des méthodes purement géométriques, certains chercheurs ont choisi de représenter les caractéristiques locales du visage sous forme de graphes. Manjunath et al. [Man92] ont proposé une méthode de détection de caractéristiques locales du visage, basée sur la décomposition en ondelettes de Gabor [Lee96]. La reconnaissance de visages est alors formulée comme un problème de mise en correspondance de graphes. L'efficacité de cette méthode a été validée sur un ensemble de données de visage de 86 sujets, contenant des variations d'expression et de pose, Un taux de reconnaissance de 90% en moyenne a été rapportée démontrant la robustesse de cette approche.

Cependant, une fois construit, le graphe topologique ne peut pas être modifié. Or, les images de 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.

Elastic Graph Matching

En se basant sur cette observation, Lades et al. [Lad93] ont proposé un graphe topologique élastique connu sous le nom de « Elastic Graph Matching ». Les auteurs utilisent pour cela une grille régulière, placée sur les images de visages. Les caractéristiques extraites sont des coefficients de Gabor. Les noeuds de la grille sont liés élastiquement. L'élasticité permet d'accommoder des distorsions d'objet et des changements de point de vue. La mise en correspondance est exécutée en deux étapes. Ce système peut s'adapter aux changements modérés de taille et d'orientation du visage.

Elastic Buch Graph Matching

La technique d'Elastic Buch Graph Matching (EBGM) [Wis97] ressemble à celle de l'Elastic Graph Matching, la différence étant que l'EBGM utilise un seul graphe pour représenter les différentes variations d'apparence de visage. Chaque nœud contient un ensemble de 40 coefficients complexes d'Ondelette de Gabor, incluant la phase et l'amplitude. Ces coefficients sont connus sous le nom de *jets*. Ainsi, la géométrie d'un objet est codée par les arêtes du graphe, alors que les nœuds (*jets*) codent les variations des niveaux de gris (ex : un « eye bunch » va contenir des jets représentant les yeux : yeux fermés, ouverts, féminins, masculins, etc.). Deux inconvénients de la méthode EBGM ont été mis en évidence [Zha97]. Premièrement, elle exige un temps de calcul supérieur aux

autres méthodes ce qui la rend plus difficile à mettre en œuvre en pratique. Et deuxièmement, seules les informations sur les positions clefs de l'image (par exemple : les yeux, le nez, la bouche) sont utilisées pour l'identification. Bien que ce soit un facteur crucial qui contribue à la robustesse de la méthode, la manière dont cette méthode gère une situation où les caractéristiques clefs sont occultées n'est pas claire.

Des améliorations ont été apportées à ce modèle par Kepenekci et al. [Kep02] : ils proposent une méthode basée sur les caractéristiques de Gabor. Au lieu de fixer le nombre de points caractéristiques du visage comme dans la méthode EBGM, ils ont utilisé un ensemble de matrices de filtres de Gabor pour parcourir les régions faciales locales. Les points caractéristiques obtenus avec la réponse fréquentielle la plus haute du filtre de Gabor sont automatiquement choisis pour être des candidats à la représentation de visage. Puisque les points caractéristiques résultants sont différents d'un visage à un autre, la possibilité de trouver des classes spécifiques de caractéristiques s'en trouve donc augmentée. En plus des valeurs de la réponse de Gabor, la position de chaque point caractéristique est enregistrée, considérant ainsi implicitement la structure spatiale du visage. Des résultats expérimentaux sur l'ensemble de la base de données ORL montrent un taux d'identification de 95.25 % avec seulement une image d'apprentissage par personne. Un deuxième test sur la base FERET [Phi00] a démontré que cette amélioration méthode est moins coûteuse en de temps de calcul que la méthode EBGM. Cependant, sa flexibilité dans la détection des points d'intérêt augmente le risque des faux appariements en raison de la possibilité de la non-existence des caractéristiques dans la zone locale considérée.

Les méthodes basées sur les caractéristiques locales sont efficaces. Cependant leurs performances dépendent essentiellement de la précision de la localisation des points caractéristiques. Cette tâche reste très difficile en pratique, plus particulièrement dans des situations où la forme et l'apparence du visage peuvent fortement changer. Par exemple, la sur-illumination peut provoquer une réflexion spéculaire sur le visage. Pour résoudre ce problème des méthodes basées sur l'apparence locale sont utilisées. Elles font l'objet de la section suivante.

2.3.2.2 Méthodes basées sur l'apparence locale (Local appearance-based methods)

Ces techniques sont utilisées de manière modulaire pour les différentes régions faciales. Un modèle global est alors défini à partir de la combinaison des différents modèles locaux. Ainsi, les différentes régions faciales ne seront plus affectées de la même manière par les différentes sources de variabilité. Par exemple, le port de lunettes de soleil change considérablement l'aspect des yeux, tandis qu'un sourire affecte plus la région de la bouche. Deux paramètres sont utilisés pour définir les régions locales du visage : la forme (rectangle, ellipse) et la taille. Les caractéristiques des régions locales sont déterminées à partir d'une analyse des valeurs de niveau gris [Che04, Tan05]. Des techniques comme les Ondelettes de Gabor [Man92, Kep02] ou de Harr [Le04] et l'analyse fractale sont aussi utilisées pour l'extraction de caractéristiques. En général, les caractéristiques à base de valeurs de gris préservent l'information de texture, tandis que les caractéristiques de Gabor sont plus robustes face aux changements d'illumination et aux transformations géométriques [Man92, Lad93].

2.3.2.2.1 ACP et ADL modulaire

Dans [Pen94], Pentland et al. ont introduit l'approche des « Modular Eigenspace ». Les régions faciales utilisées sont : la totalité du visage, les yeux et le nez. Une ACP est appliquée sur chacune de ces régions et les résultats de classifications obtenus sont agrégés. La bouche étant trop sensible à des changements d'expression faciale, sa prise en compte engendre une baisse du taux de reconnaissance. Cette approche peut être qualifiée d'hybride, puisqu'elle utilise à la fois des caractéristiques globales et locales. Les auteurs ont montré qu'elle est plus efficace que les techniques globales ou strictement locales appliquées séparément.

Comme expliqué ci-dessus, les méthodes basées ADL échouent si un seul échantillon par classe est utilisé. Chen et al. [Che04] ont proposé d'appliquer la méthode ADL sur un exemple de petite taille. Ils ont partitionné chaque image de visage en un ensemble d'images de même dimension (figure 2.5).



Figure 2.7. Image de visage partitionnée en imagerie[Che04].

Pour chaque classe, des échantillons composés de toutes les imagerie partitionnées sont produits pour l'apprentissage. Enfin, la méthode ADLF (Analyse Discriminante Linéaire Fisher) est appliquée au échantillons nouvellement produits. Cette méthode a été évaluée sur un sous-ensemble de données FERET contenant 200 personnes avec une imagerie d'apprentissage par personne : un taux d'identification de 86.5 % a été atteint.

Dans [Pri05], Price et Gee ont introduit une technique modulaire basée sur une variante de l'ADL. Les régions sélectionnées sont : la région faciale dans son ensemble, une bande faciale (de même largeur que la région faciale) s'étalant du front jusqu'au-dessous du nez, et une bande faciale contenant les yeux. Les résultats expérimentaux montrent que cette approche est plus performante que les techniques des *eigenfaces* et des *fisherfaces*, elle est notamment robuste aux changements dans les conditions d'illumination du visage, d'expression faciale et d'occultation.

2.3.2.2.2 Les approches probabilistes Locales modulaires

Martinez et al [Mar02] présentent une approche probabiliste locale pour la reconnaissance de visages avec des occultations partielles et des variations d'expression. Ils utilisent un seul exemple d'apprentissage par classe. Chaque visage est divisé en six régions locales sous forme d'ellipses. Les régions ayant les mêmes positions pour chaque visage sont groupées séparément dans un sous-espace. Pour une représentation plus compacte et efficace, chaque sous-espace de visage est à nouveau transformé dans un *eigenspace*, où la distribution est évaluée par la moyenne d'un modèle de mélange de Gaussiennes en utilisant l'algorithme EM. Dans l'étape d'identification, les images de test sont aussi divisées en six régions locales et sont projetées respectivement dans l'*eigenspace* calculé.

Une approche probabiliste plutôt qu'une approche de vote est utilisée pour mesurer la similitude d'un appariement donné. Les tests sur un jeu de 2600 images ont montré que l'approche probabiliste locale ne réduit pas la précision même pour 16 des visages partiellement occultés. Cependant, le mélange de Gaussiennes utilisé dans cette méthode est de nature paramétrique, il dépend de l'hypothèse que la distribution puisse être fidèlement représentée avec les exemples donnés.

Bien qu'un certain nombre d'exemples soit synthétiquement produit comme décrit ci-dessus, les coûts de calcul et de stockage avec la procédure de production des exemples virtuels peuvent être très élevés (par exemple 6615 exemples par individu dans [Mar02]) quand la base de données de visage est très grande.

Tan et al. [Tan05] ont étendu l'approche probabiliste locale en proposant une méthode alternative pour représenter le sous-espace de visages en utilisant la carte SOM (Self-Organizing Map) [Koh97]). Plus spécifiquement, chaque image I de visage est d'abord divisée en M différents sous blocs locaux $R_i \Big|_{i=1}^M$, puis un réseau SOM est entraîné en utilisant tous les sous-blocs obtenus à partir de toutes les images d'apprentissage disponibles, sans tenir compte des classes. Chaque sous-bloc R_i de la même image I de visage est ensuite mappé avec les unités qui correspondent le mieux (BMU : best matching units) selon la stratégie suivante : le plus proche voisin, dans la SOM 2D de l'espace topologique, est défini comme un vecteur de position $l_i = \{x_i, y_i\}$. Tous les vecteurs de position du même visage sont groupés en un ensemble, c'est-à-dire $I = \{l_i\}_{i=1}^M = \{x_i, y_i\}_{i=1}^M$, appelé la représentation "SOM-FACE" du visage (figure 2.6b). Cette représentation possède plusieurs avantages. Des défauts possibles, comme le bruit dans l'image de visage originale, peuvent être éliminés durant le processus d'entraînement de la SOM. Cette méthode est une façon compacte de représenter le visage.

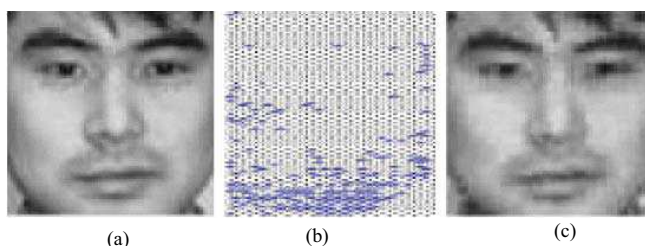


Figure 2.8. Exemple : (a) une image du visage originale, (b) sa projection (SOM-Face) et (c) image reconstruite [Tan05].

À la différence d'autres méthodes comme l'ACP, cette représentation est intuitivement compréhensible. En effet, chaque élément d'un SOM-VISAGE a une signification physique donnée par le vecteur de poids stocké dans le nœud correspondant de la carte SOM, qui peut être interprété comme une pièce (patch) faciale locale dans l'espace d'entrée.

2.3.2.2.3 Les Modèles de Markov Caché modulaire

Les méthodes mentionnées ci-dessus ne considèrent pas explicitement le rapport entre les caractéristiques locales. Une autre façon intéressante d'incorporer des informations globales est la méthode basée sur les Modèles de Markov Caché (MMC). Plutôt que de traiter l'image de visage comme un graphe topologique statique où les caractéristiques locales représentent les nœuds, la méthode basée sur les MMC caractérise le modèle de visage comme un processus aléatoire dynamique avec un ensemble de paramètres.

Samaria et al. [Sam93] ont illustré l'utilité des techniques MMC dans l'identification de visages. Dans leur méthode, un modèle de visage est divisé en cinq régions qui se chevauchent, incluant le front, les yeux, le nez, la bouche et le menton. La technique MMC est introduite en considérant chaque région comme un état caché d'un modèle MMC. Un modèle de visage est alors considéré comme une séquence d'observation constituée de cinq états, dont chacun peut être modélisé par une distribution gaussienne multi-variables, et les transitions probabilistes entre les états peuvent être apprises à partir des frontières entre les régions. Après que le MMC soit entraîné, un visage peut être reconnu en calculant la probabilité de sortie de son ordre d'observation. L'inconvénient de cette méthode est qu'elle est trop exigeante en termes d'échantillons d'apprentissage afin d'assurer la fiabilité de l'estimation des paramètres.

Le et al. [Le04] ont présenté une méthode pour rendre la technique MMC applicable dans le cas d'un exemple d'apprentissage. Deux facteurs ont contribué à la faisabilité et à l'efficacité de leur méthode. D'abord, ils ont produit une grande collection de vecteurs d'observation de chaque image, dans les deux directions verticale et horizontale, agrandissant ainsi l'ensemble d'apprentissage. Deuxièmement, l'Ondelette de Haar a été appliquée à l'image pour diminuer la dimension des vecteurs d'observation et améliorer la robustesse. Leurs résultats expérimentaux évalués sur la base de données de visage frontal

"AR" montrent que la méthode proposée surpasse les approches PCA, LDA et l'analyse de caractéristiques locales (LFA local feature analysis [Pen96]).

2.3.2.3 Discussion

Nous avons passé en revue les méthodes locales traitant le problème de la reconnaissance de visages dans le cas d'un seul exemple d'apprentissage. Nous les avons classifié en deux catégories principales : les méthodes à base de caractéristiques et les méthodes basées sur l'apparence locale. En réalité, ces deux catégories ne sont pas si différentes car les régions locales sont constituées d'un ensemble de pixels parmi lesquels des points caractéristiques intéressants peuvent être détectés.

Bien que les méthodes locales aient prouvé leur efficacité dans le cas d'un seul exemple d'apprentissage, plusieurs problèmes restent toujours non résolus, comme par exemple le choix des caractéristiques locales qui n'est pas du tout évident. De plus, les méthodes locales sont robustes uniquement vis-à-vis d'un nombre restreint de facteurs de variations. Ainsi, la méthode EGBM est robuste à certains changements d'expression, d'illumination et de pose, mais pas aux occultations ; en revanche la méthode probabiliste locale et la méthode SOM-Face sont robustes aux variations d'expression et aux occultations, mais pas vis-à-vis des changements de pose. Une voie possible pour améliorer la robustesse d'un système de reconnaissance de visages peut résider dans les méthodes hybrides qui combinent différentes techniques de reconnaissance.

2.3.3 Méthodes Hybrides

Les méthodes hybrides sont des approches qui combinent les caractéristiques holistiques et locales afin d'améliorer les performances de la reconnaissance de visages. En effet, les caractéristiques locales et les caractéristiques globales ont des propriétés tout à fait différentes. On peut espérer pouvoir exploiter leur complémentarité pour améliorer la classification. Le tableau 2.2 récapitule qualitativement la différence entre les deux types de caractéristiques [Tan06].

Facteurs de variations	Caractéristiques locales	Caractéristiques globales
Illuminations [Hal99]	Très sensible	Sensible
Expressions [Mar02][Tan05]	Pas sensible	Sensible
Pose [Hei01]	Sensible	Très sensible
Bruit [Cos02]	Très sensible	Sensible
Occlusion [Mar02][Tan05]	Pas sensible	Très sensible

Tableau 2. 3. Comparaison des méthodes basées sur les caractéristiques locales ou globales.

Nous pouvons voir que les caractéristiques locales et globales réagissant différemment sensibles aux facteurs de variation. Par exemple, les changements d'illumination peuvent avoir plus d'influence sur les caractéristiques locales, tandis que les changements d'expression ont plus d'impact sur les caractéristiques holistiques. Ainsi, les méthodes hybrides peuvent constituer une approche efficace pour réduire la complexité des classifieurs et améliorer leur capacité de généralisation. Malgré ces avantages, les travaux de recherche dans ce domaine restent relativement peu nombreux, probablement en raison des difficultés mentionnées ci-dessus, tandis que les méthodes hybrides classiques (c'est-à-dire qui utilisent plusieurs exemples d'apprentissage par personne), comme les modèles d'apparence flexibles [Lan95] ou le LFA hybride [Pen96], ne sont pas généralement appropriées pour traiter le cas d'un seul exemple d'apprentissage.

2.3.3.1 Les Modèles Actifs d'Apparence

[Lan95] a utilisé la méthode dite « Modèle Actif d'Apparence » (MAA) pour la reconnaissance de visages. Le MAA présenté par [Coo01] constitue un outil d'extraction des caractéristiques (signatures) représentant, à la fois, la forme et la texture des visages. La base d'apprentissage de 400 images est annotée très précisément et manuellement par 68 points caractéristiques pour modéliser la forme du visage. Approximativement 10,000 exemples de régions faciales ont été utilisés. Le processus de classification est illustré par la figure 2.7.

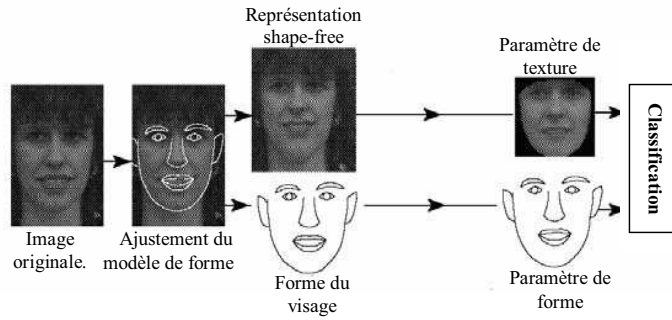


Figure 2. 9. Processus de reconnaissance de visages basé sur les Modèles Actifs d'Apparence [Lan95].

Chaque exemple est représenté par un vecteur de forme x contenant les caractéristiques du visage. À partir des vecteurs de forme de la base d'apprentissage, une ACP est appliquée afin d'extraire les modes de variations principaux du vecteur de forme x , qui peuvent ensuite être estimés par le vecteur \bar{x} , tel que :

$$\bar{x} = \bar{x} + P_f b_f \quad (2.16)$$

Où \bar{x} est la forme moyenne de la base d'apprentissage. Le vecteur $b_f = P_f^T x$ correspond à la projection de x dans P_f , et constitue le vecteur des paramètres de forme. Les textures des visages (valeurs des pixels en niveaux de gris) sont alors normalisées au sens de leur forme. On obtient un visage dit shape-free. Les textures sont déformées selon la forme moyenne (à l'aide d'un algorithme de triangulation par exemple). Pour chaque visage, on obtient ainsi un vecteur de texture g , indépendant de la forme de visage. Une ACP est appliquée sur ces vecteurs de texture. De même, le vecteur g peut donc être approximé par :

$$\bar{g} = \bar{g} + P_g b_g \quad (2.17)$$

Où b_g est le vecteur de paramètres de texture associé au visage. Chaque image est donc caractérisée par ses vecteurs b_f et b_g , qui sont corrélés. Afin de les décoller, on concatène les vecteurs obtenus. On obtient alors le modèle combiné suivant :

$$\bar{x} = \bar{x} + Q_f c \quad (2.18)$$

$$\bar{g} = \bar{g} + Q_g c \quad (2.19)$$

Où c est le vecteur d'apparence contrôlant à la fois la forme et la texture du modèle, et Q_f et Q_g sont respectivement les matrices de projection de c dans les espaces de variations de forme et de texture.

Pour un vecteur de paramètre c fixé, on peut synthétiser l'image de visage associée. Pour cela, on génère le visage shape-free correspondant, puis on procède à sa déformation en utilisant les points de contrôle du vecteur x . Lorsqu'un visage requête doit être reconnu, le but est de déterminer le paramètre c optimal. Il s'agit de la valeur c minimisant l'erreur entre l'image originale et l'image synthétisée. Cette technique présente le désavantage de reposer sur une procédure d'optimisation coûteuse, instable, instable et dépendant de nombreux paramètres.

Enfin, nous pouvons noter que, d'un certain point de vue, les méthodes locales peuvent être considérées comme des méthodes hybrides, puisque les informations globales sont utilisées d'une façon indirecte dans l'algorithme de reconnaissance. Dans la méthode probabiliste locale [Mar02], par exemple, de nouveaux exemples d'apprentissage pour chaque personne sont d'abord produits avec la méthode holistique, la méthode locale est ensuite utilisée pour l'identification.

Dans le tableau (2.3) résume les performances des différentes méthodes utilisant un seul exemple pour l'apprentissage.

Méthodes	Base de données	Nbr personnes	Nbr images	RR (%)	Variations
(PC) ² A [Wu02]	FERET	200	200	83.5	N/A
ACP2D [Yan04]	AR	100	600	74.8	Expression
ACP Discriminant [Wan05]	FERET	256	914	72	Expression
ADLF modulaire [Che04]	FERET	200	200	86.5	Expression
Probabiliste Locales	AR	200	200	82.3	
modulaire	AR	100	400	71	Occlusion
[Mar02]					
---	AR	100	600	93.7	Expression
SOMface [Tan05]	AR	100	400	76	Occlusion
----	AR	120	1440	89.8	Expression,
1D-DHMM [Le04]					Illumination
EBGM [Wis97]	FERET	1196	1196	95%	Expression,
					Illumination

Tableau 2. 4. Comparaison des performances des méthodes utilisant un seul exemple pour l'apprentissage

2.4 Techniques 3D de reconnaissance de visages

Nous avons passé en revue les méthodes de reconnaissance 2D de visages. Malgré les avancées réalisées ces dernières années, les techniques de reconnaissance 2D de visages robustes aux différents facteurs de variabilité (éclairage, pose, occultation) sont loin d'être développées. La reconnaissance 3D de visages constitue une alternative prometteuse pour surmonter ces problèmes, surtout depuis l'apparition de dispositifs d'acquisition 3D performant. L'avantage principal des approches basées modèle 3D réside dans le fait que le modèle 3D conserve toutes les informations sur la géométrie de visage, ce qui permet d'avoir une représentation réelle de ce dernier. Dans cette section, après avoir rapidement évoqué les systèmes d'acquisition 3D, nous présenterons les travaux récents sur la reconnaissance 3D de visages.

2.4.1 Systèmes d'acquisition 3D

Généralement le modèle du visage est représenté par des images 2.5D et 3D (voir figure 2.8). L'image 2.5D (image de profondeur) correspond à une représentation bidimensionnelle d'un ensemble de points 3D (x,y,z) où chaque pixel dans le plan X-Y stocke la valeur de sa profondeur z . On peut assimiler une image 2.5D à une image en niveau de gris où les pixels noirs correspondent au fond tandis que les pixels blancs représentent les points de surface les plus proches de la caméra. Par ailleurs, la méthode la plus simple pour représenter un visage 3D est le maillage polygonal 3D, ce dernier correspond à une liste de points connectés par des arêtes (polygones). Il existe plusieurs techniques pour construire un maillage 3D, les plus utilisées combinent des images 2.5D ou bien exploitent des systèmes d'acquisition 3D tel que le scanner 3D.

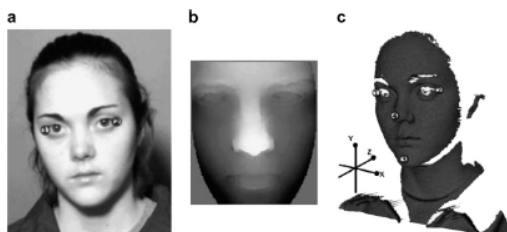


Figure 2. 10. (a) Image texture (b) Image 2.5D (c) Image 3D

Différents systèmes d'acquisition sont utilisés pour la reconstruction 3D de visages [Bou06]. Ces derniers peuvent être décomposés, en fonction des capteurs de vision qu'ils utilisent, en deux grandes classes : la vision dite active et la vision dite passive.

- Nous appelons *vision active* tout système de vision utilisant un capteur actif qui permet d'illuminer les objets de la scène d'une manière particulière en projetant sur leur surface des motifs structurés (un spot, un plan lumineux, une grille, etc.), ceci permet d'en extraire des informations 3D indépendamment des conditions propres d'éclairage de la scène. Ainsi, l'objectif de la vision active est de réduire la complexité des traitements à effectuer pour reconstruire la scène. En effet, les motifs projetés sur la scène sont facilement identifiables [Hu89] à l'inverse des primitives « naturelles » tel que les contours. Le scanner 3D, par exemple, est un système de vision active en lumière structurée qui est capable de scanner les formes 3D via un laser et les textures (couleur de la peau) via une caméra simultanément.
- En *Vision passive*, la vision binoculaire ou vision stéréoscopique (utilisation de deux caméras fixes). La vision stéréoscopique sera détaillée au chapitre 5.

Les techniques de reconnaissance 3D de visages peuvent être regroupées en trois catégories principales : approches basées modèle, approches 3D, et approches multimodales 2D + 3D.

2.4.2 Approches modèle

Ces approches construisent, à partir des points 3D, des modèles de visages qu'elles utilisent par la suite pour la reconnaissance. Blanz et al.[Bla03] [Bla02] ont proposé une méthode basée sur un modèle 3D « *morphable* » du visage. L'ensemble des visages est représenté par un espace vectoriel [Vet97]. La base de données contient 100 visages d'hommes et 100 visages de femmes acquis avec un *scanner laser CyberwareTM 3030PS*. Les points 3D des modèles de visages générés sont représentés par leurs coordonnées cylindriques définies par rapport à un axe vertical. Pour chaque visage de référence, les coordonnées et les valeurs de texture de tous les *sommets* ($n = 75 \text{ } 972$) sont regroupées pour former deux vecteurs : un vecteur de forme et un vecteur de texture. Une fois le modèle générique créé, l'étape suivante consiste à l'ajuster sur l'image 2D à partir des paramètres de forme et de texture. La synthèse d'image permet de rendre les nouvelles positions projetées des sommets du modèle 3D, à l'aide l'illumination et la couleur

extraites. Enfin, l'étape de reconnaissance est réalisée en mesurant la distance de Mahalanobis [Dud01] entre la forme et les paramètres de texture des modèles dans la galerie et le modèle d'ajustement. L'identification a été évaluée sur deux bases de données d'images à accès libre : CMU-PIE [Sim03] et FERET [Phi00]. Un taux de reconnaissance de 95 % sur l'ensemble de données CMU-PIE et 95.9 % sur l'ensemble de données FERET a été obtenu.

2.4.3 Approches 3D

Elles sont subdivisées en deux catégories : les approches basées surface qui utilisent la géométrie de la surface du visage et les approches holistiques 3D.

2.4.3.1 Approches surface

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 [Bes92]. Il consiste en une optimisation alternée d'appariements et de transformations. Ainsi, à partir d'une transformation initiale, les deux étapes suivantes sont répétées :

- *mise en correspondance* (plus proche voisin) : on apparie chaque primitive du modèle transformé avec la primitive la plus proche dans la scène.
- *recalage* : la transformation (translation + rotation) est généralement calculée aux sens des moindres carrés, surtout si l'on travaille avec des points. Si l'on possède une information d'incertitude, on peut l'utiliser dans les étapes terminales pour affiner la solution.

Une approche basée sur l'algorithme ICP est donnée par Cook et al [Coo04]. Ils utilisent l'algorithme ICP pour établir la correspondance entre les surfaces 3D afin de compenser des erreurs dues à la nature non rigide des visages. En effet, les visages sont comparés en employant un modèle statistique, en l'occurrence un Mélange de Modèles Gaussiens (MMG). La distribution des erreurs est alors paramétrée. Les auteurs ont réalisé des expériences sur la base de données 3D RMA [Beu00], une base créée au sein du *Signal and Image Center du Royal Military Academy* (Bruxelles - Belgique). Un dispositif de vision active (*structured light-based*) est utilisé pour l'acquisition. Cette base contient les images

de 120 individus, capturées en deux sessions, sous formes de points 3D non maillés. Un taux d'identification de 97.33% a été obtenu. Irfanoglu et al. [Irf04] proposent une autre approche qui utilise l'algorithme ICP pour mettre en correspondance les points caractéristiques de visages. Ils ont développé une méthode pour obtenir un appariement dense de points du maillage contenant les points présents sur tous les visages, de sorte que l'alignement de visage soit trivialement obtenu. Une fois que l'appariement dense est établi, la distance PSD "Point Set Distance", qui représente une approximation discrète du volume entre les surfaces faciales, est employée pour calculer la distance entre deux nuages différents de points. Dans leur expérimentation, les auteurs ont testé l'algorithme sur la base de données 3D RMA, ils ont relevé un taux d'identification de 96,66%.

Quoique l'ICP soit un outil puissant pour évaluer la similitude entre deux visages, les méthodes basées ICP ne sont pas capables de traiter le problème de changement d'expression, car elles considèrent le visage 3D comme un objet rigide, ce qui n'est pas le cas dans la réalité. Pour résoudre ce problème, plusieurs approches ont été développées. Moreno et al. [Mor03] proposent de segmenter le visage en plusieurs régions en utilisant la courbure gaussienne. Ainsi, chaque modèle facial 3D est subdivisé en 86 régions différentes grâce à un algorithme de segmentation qui exploite les signes de la médiane et des courbures gaussiennes pour isoler les régions ayant des courbures significatives. Cet espace de caractéristiques est ensuite réduit pour augmenter l'efficacité de la méthode. Enfin, un vecteur de caractéristiques est créé pour chaque sujet. Les expérimentations ont été menées sur un ensemble de 420 modèles 3D de visages appartenant à 60 sujets, comprenant des images avec des variations de lumière, de rotation et d'expressions de visages. Les auteurs ont obtenu un taux de reconnaissance de 78 %.

Le processus de segmentation peut aussi être utilisé pour traiter la reconnaissance de visage comme un problème de reconnaissance d'objets non rigides, et ainsi améliorer la robustesse de la reconnaissance face aux variations d'expressions. Chua et al. [Chu00] ont remarqué que des régions du visage, comme le nez, l'orbite de l'œil et le front subissent beaucoup moins les déformations en cas de changements d'expressions. Ils identifient les régions faciales "rigides" en utilisant une comparaison "deux par deux" entre les signatures de points [Chu97] pour différentes expressions de visage de la même personne. Ainsi, seulement les parties rigides sont stockées dans une base de données indexée. Les modèles de la base sont classés en fonction de leur similitude. Leur expérimentation a montré un

taux de reconnaissance de 100 % sur une base de données de 60 sujets et 4 variations d'expressions de visage. Enfin, [Ben06] propose d'introduire une nouvelle métrique orientée région dans l'algorithme de l'ICP. Il s'agit de segmenter un visage en plusieurs régions ayant des influences plus ou moins importantes sur la déformation de la forme 3D du visage. Le calcul de la distance (similarité) est pondéré en fonction de ces régions en donnant plus de poids aux régions statiques qu'aux régions mimiques du visage. Un taux de reconnaissance de 97.56 % a été obtenu.

2.4.3.2 Approches holistiques 3D

Les techniques holistiques comme l'ACP ont été largement utilisées dans la reconnaissance faciale 2D. Plus récemment, ces techniques ont été aussi étendues aux données 3D de visage. Ainsi, Heshner et al. [Hes03] ont appliqué l'ACP sur une base de données contenant 222 images de profondeur (37 personnes), avec six expressions faciales différentes. Ils ont obtenu un taux de reconnaissance de 83%. Heseltine et al. [Hes04a] [Hes04b] ont développé deux approches qui appliquent l'ACP sur des représentations tridimensionnelles du visage. La première méthode [Hes04a] utilise l'approche basée ACP « eigensurface » sur des modèles 3D maillés de visage de la base de données de l'Université de York (The 3D Face Database, 2003). Elle arrive à un taux de reconnaissance de 87%. La deuxième approche est une extension de la méthode Ficherface 2D de Belhumeur [Bel97] aux données surfaciques de visage 3D. Les tests ont été effectués sur une base de données contenant 1770 modèles 3D maillés de 280 sujets ayant différentes poses et expressions faciales. Le taux de reconnaissance obtenu est de 88%.

Les techniques basées ACP ont également été combinées avec d'autres méthodes de classification, comme le modèle caché de Markov (EHMM) puis appliquées à la reconnaissance 3D de visages [Tsa04]. Enfin, d'autres approches basées sur l'Analyse Discriminante Linéaire [Gok05] ou l'Analyse des Composantes Indépendantes [Kim03] ont aussi été développées pour l'analyse des données 3D de visages.

2.4.3.3 Approche géométrique ou locale 3D

Par rapport aux approches « holistiques », les techniques d'identification 3D du visage basées sur les caractéristiques faciales locales de type géométriques restent relativement peu développées, alors que dans le même temps des approches locales d'identification 2D,

telles que l'« Elastic Bunch Graph Matching (EBGM) [Wis97] » et l'analyse locale de caractéristiques, connaissent un grand succès auprès des fournisseurs de données (Recognition Vendors Test 2002 [Phi03]). Par ailleurs, les combinaisons des techniques de reconnaissance 2D et 3D sont de plus en plus étudiées. Ainsi la combinaison des approches EBGM 2D et EBGM 3D (qui correspond à une extension de la méthode EBGM 2D aux images de profondeur de visage) a donné lieu à une technique d'identification du visage 2D+3D très efficace évaluée sur la base FRGC [Hus05]. Cependant, l'évaluation de cette approche hybride a démontré que les performances du EBGM 2D surpassent de manière significative celles du EBGM 3D. Elle a aussi mis en avant le besoin de comprendre et d'utiliser les propriétés discriminantes des caractéristiques locales du visage afin de concevoir des techniques efficaces de reconnaissance 3D de visage.

Dans [Lee05] les auteurs proposent un système de reconnaissance de visage 3D utilisant deux dispositifs différents d'acquisition de données 3D : un scanner laser 3D et une lumière structurée. A partir des mesures 3D, huit points caractéristiques du visage (géométriquement invariables) sont extraits puis utilisés pour calculer un vecteur caractéristique comprenant les distances et les angles entre ces points. Pour la phase de reconnaissance, les auteurs ont implémenté deux algorithmes différents : le premier basé profondeur (depth-based DP) et le second utilise les SVM (feature-based SVM). Les résultats expérimentaux (sur une base de 20 personnes) ont donné un taux de reconnaissance de 95% pour la DP et de 96% pour la SVM.

2.4.4 Approches 3D+2D

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. Récemment, plusieurs approches basées sur ce principe ont été développées. Bronstein et al. [Bro03] ont proposé une méthode qui considère le visage comme une surface isométrique. Ils se sont inspirés des travaux de Schwart et al. [Sch89], les premiers à employer la graduation multidimensionnelle (MDS) pour projeter des surfaces corticales du cerveau sur un plan. Zigelman et al. [Zig02] et Grossman et al. [Gro02] ont appliqué ce principe aux problèmes de mapping de texture et de l'aplatissement «flattening» de cortex basé sur les voxels. Une généralisation de cette approche pour l'identification d'objet a été présentée dans les travaux d'Elad et de Kimmel [Ela01]. Ils ont développé un algorithme efficace pour

construire une signature pour les surfaces isométriques. Cette méthode est désignée sous le nom des formes canoniques « bending-invariables ». Elle calcule des distances géodésiques entre les points prélevés sur la surface faciale. En plus de l'image 2D de texture, les auteurs ont utilisé une lumière structurée pour acquérir des images de profondeur (géométrie) des visages. L'image est convertie en une surface triangulée (maillage triangulaire) puis lissée en utilisant les splines. La taille de la surface est ensuite réduite approximativement à 2000-2500 sommets. Enfin, la forme canonique bending-invariable du visage est calculée puis alignée en utilisant la méthode décrite dans [Bro03]. Les auteurs appliquent la technique eigenforms sur les vecteurs canoniques pour la reconnaissance de visage. Cette méthode a été testée sur une base de données contenant 220 images de 30 personnes (27 vraies personnes et 3 mannequins) ; un taux d'identification de 100 % est rapporté.

Chang et al. [Cha03] ont mené une étude expérimentale poussée qui leur a permis de comparer et de combiner des données 2D et 3D pour la reconnaissance faciale. Un total de 275 sujets a été sollicité pour l'acquisition des visages. 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. Il est clair que durant le processus de décision, certaines métriques sont plus ou moins appropriées à une modalité donnée. Dans leur expérimentation, les auteurs ont utilisé la distance de Mahalanobis pour réaliser la mise en correspondance (matching). La métrique utilisée pour la fusion de données combine les distances d'appariements trouvées séparément dans chaque espace. Il existe plusieurs méthodes pour combiner des métriques différentes afin d'obtenir la meilleure règle de décision, telle que la somme, la multiplication, le minimum, la moyenne, etc.

La décision multi-modale développée par les auteurs peut être résumée comme suit : dans un premier temps, les images d'entrée 2D (2D probe) et 3D sont appariées avec les images des galeries 2D et 3D respectivement. Ceci permet d'obtenir deux ensembles de N distances dans deux espaces différents, l'espace facial 2D et l'espace facial 3D. N est la taille de la galerie d'images. Les distances 2D et 3D sont additionnées, et l'image qui donne la plus petite somme est sélectionnée. Plusieurs tests ont été menés pour évaluer cette méthode. Une première expérimentation utilise des images anciennes des différents sujets. Une autre expérience emploie un ensemble plus grand constitué de 676 images de test prises lors de différentes sessions d'acquisition espacées dans le temps. Le taux

d'identification obtenu pour les deux expériences est approximativement égal à : 99% pour l'approche multi-modale 3D + 2D, 94% pour l'approche 3D seule, et 89% pour l'approche 2D seule. Le résultat de l'approche multi-modale est obtenu en utilisant une somme pondérée des distances dans les espaces de visage 3D et 2D. Cette étude a démontré, grâce à l'utilisation de l'ACP sur les images 2D et 3D, que les données faciales 3D fournissent des performances biométriques bien meilleures que les données faciales 2D. Par ailleurs, les auteurs ont démontré aussi que la combinaison des données faciales 2D et 3D permet d'augmenter d'une manière significative les performances de la reconnaissance.

Dans [Wan02] les auteurs ont développé une approche 2D+3D combinant deux méthodes de localisation. Une méthode qui utilise un filtre de Gabor [Wis97] pour localiser les points d'intérêt 2D, et une deuxième méthode qui utilise l'algorithme de [Chu97] pour extraire les points de signature 3D (voir figure 2.9).



Figure 2. 11. Choix de dix points : points 2D marqués par le " o " et points 3D marqués par "x"
[Wan02].

Les points 3D et les points de texture 2D ainsi obtenus sont projetés chacun dans leur propre sous-espace, en utilisant l'ACP, pour générer respectivement des vecteurs de forme et de texture. Ces vecteurs sont regroupés en un seul vecteur qui est normalisé afin de représenter l'image faciale. Le processus d'identification est basé sur une fonction de similitude utilisant la méthode SVM. Les expérimentations, faisant participer 50 personnes ayant des expressions faciales différentes pour des points de vue différents, ont démontré l'efficacité de l'algorithme. Un taux reconnaissance de 90 % a été obtenu.

2.4.5 Conclusion

Dans ce chapitre, nous avons passé en revue les principales techniques de détection et de reconnaissance de visages. Le problème n'est pas du même niveau de complexité selon que l'on cherche à détecter un visage dans une image qui en contient un seul de taille à peu près connue, ou que l'on se place dans l'optique d'une détection de multiples visages de

tailles variables. Dans le cadre de ce travail nous nous limiterons à la détection d'un visage unique dans une image, contexte du projet IV² et les conclusions que nous exposons se situent dans cette optique. Parmi les différentes caractéristiques invariantes du visage utilisées pour la détection, la couleur de la peau paraît parmi les plus prometteuses. En effet, les méthodes basées sur la couleur de la peau sont robustes aux occultations ainsi qu'aux variations de la pose, de l'échelle, et de l'expression. Ces méthodes restent cependant peu robustes au changement d'éclairage. Toutefois, comme il est mentionné dans [Kak06], des techniques comme le classifieur de Bayes ou bien les réseaux de neurones peuvent conduire à des performances satisfaisantes lorsqu'elles sont appliquées sur des bases de données de petites tailles. Nous avons montré que des approches qui utilisent des prétraitements préalables, comme par exemples la correction de couleur « color constanty », permettent d'améliorer les performances de la détection et constituent une solution intéressante au problème de l'illumination. Dans le cadre de cette thèse, nous proposons une approche de détection de visage basée sur une segmentation couleur adaptative qui traite le problème de l'illumination.

Par ailleurs, nous avons classifié les techniques de reconnaissance 2D de visages en trois grandes catégories : les approches globales, les approches locales et les approches hybrides. Nous avons souligné que l'ACP permet de réaliser un processus de reconnaissance même si l'on ne possède qu'une seule image d'une personne, ce qui n'est pas possible avec des méthodes telles que l'analyse discriminante linéaire ou les « Support Vector Machine ». Toutefois, nous avons soulevé la difficulté de l'ACP, et de ses différentes extensions, à gérer les occultations et les variations de poses, d'illumination et d'expressions. Les approches locales ont prouvé leur efficacité dans le cas d'un seul exemple d'apprentissage, mais elles souffrent d'un manque de robustesse (la méthode EGBM, par exemple, n'est pas robuste face aux occultations). Enfin, les approches hybrides semblent intéressantes pour résoudre le problème de la robustesse, cependant elles ne sont pas appropriées pour l'instant pour traiter le cas d'un seul exemple d'apprentissage (tel que les modèles d'apparence flexibles [Lan95] ou le LFA hybride [Pen96]).

La reconnaissance 3D de visages constitue une alternative prometteuse pour résoudre les problèmes de robustesse que connaît la reconnaissance 2D. Cependant, elle n'a pas encore atteint une certaine maturité, à cause notamment de la lourdeur du processus d'acquisition,

et de la non-disponibilité de grandes bases de données de visages 3D, à accès libre, afin de tester et d'évaluer les techniques élaborées. Les approches de reconnaissance 3D de visage développées jusqu'à présents sont soit des extensions d'approches 2D, comme l'ACP qui en outre donne des résultats plus précis en 2D qu'en 3D, soit des adaptations d'approches 3D de reconnaissance de formes rigides, tel que l'ICP qui donnent des taux d'identification plus intéressants, mais dont l'inconvénient majeur reste l'initialisation du recalage des surfaces faciales. Par ailleurs, les algorithmes de reconnaissance 3D de visage basés sur les caractéristiques faciales locales ne sont pas fondés sur aucune étude de la structure du visage. Les points d'intérêt du visage sont plutôt choisis arbitrairement [Mor03, Lee05, Wan02] ou bien correspondent à une extension de caractéristiques locales 2D [Hus05]. En comparaison avec les images faciales 2D qui donnent des informations sur la texture faciale, les modèles faciaux 3D fournissent des informations riches sur la structure même du visage, ce qui rend le problème de reconnaissance 3D de visage fondamentalement différent de celui de la reconnaissance 2D.

Enfin, nous avons souligné l'avantage des techniques multi modales qui combinent des données 2D et 3D de visages pour améliorer les performances et la robustesse de la reconnaissance. Nous avons présenté quelques travaux qui ont clairement rapporté des taux de reconnaissance nettement supérieurs à ceux des techniques de reconnaissance 2D et 3D isolées.

Dans la suite de ce travail nous allons présenter les études que nous avons faites dans le cadre de la vérification d'identité. Dans ce contexte suite aux travaux que nous avons présenté dans ce chapitre 2, nous avons décidé d'utiliser une stratégie de reconnaissance basée sur la combinaison 2D/3D. Dans le chapitre 4 nous détaillerons le traitement réalisé permettant d'extraire un visage, puis des régions de ce visage à travers une implémentation de l'algorithme de Ligne de Partage des Eaux sur des images couleur. Dans le chapitre 5, nous mettons en avant l'intérêt de l'ACP modulaire sur les images 2D, puis nous proposons une méthode de vérification d'identité à partir de données 3D, enfin nous explorerons les possibilités de combinaison de ces deux types d'information. Mais avant nous allons présenter les données utilisées pour nos évaluations. C'est ce qui fera l'objet du chapitre 3.

Chapitre 3

Bases de données utilisées

3.1 Généralité sur les bases de données existantes

Plusieurs bases de données contenant des informations qui permettent l'évaluation des systèmes de reconnaissance de visages sont disponibles sur le marché. Toutefois, ces bases de données sont généralement adaptées aux besoins de quelques algorithmes spécifiques de reconnaissance. Le tableau (3.1) récapitule les principales caractéristiques de ces bases [Li05]. Il résume, pour chacune d'elle, les conditions d'acquisition des images de visages (changements d'illumination, de pose, d'expression faciale), la disponibilité ou non des données 3D et aussi nombre de sessions d'acquisition. Par ailleurs, plusieurs protocoles expérimentaux ont été élaborés afin d'évaluer le nombre important d'algorithmes de reconnaissance proposés ces dernières années. Des efforts particuliers de standardisation ont été aussi déployés pour définir des protocoles d'évaluation standardisés. Ainsi, la série d'évaluations FERET [Phi89b] menées par le « National Institute of Standards and Technology (NIST) » a permis la comparaison entre neuf systèmes de reconnaissance de visages proposés par des institutions et des entreprises dans le contexte de l'identification et de la vérification. D'autres évaluations ont été effectuées par la suite, notamment celles du « Face Recognition Vendor Test (FRVT) » [Phi03].

Base de données	Nombre de personnes	Pose	Illumination	Facial Expression	Temps	Données 3D
AR	116	1	4	4	2	non
BANCA	208	1	++	1	12	non
CAS-PEAL	66-1040	21	9-15	6	2	non
CMU Hyper	54	1	4	1	1-5	non
CMU PIE	68	13	43	3	1	non
Equinox IR	91	1	3	3	1	non
FERET	1199	9-20	2	2	2	non
Havard RL	10	1	77-84	1	1	non
KFDB	1000	7	16	5	1	non
MIT	15	3	3	1	1	non
MPI	200	3	3	1	1	oui
NDHID	300	1	3	2	10/13	oui
NIST MID	1573	2	1	--	1	non
ORL	10	1	--	--	--	non
UMIST	20	--	1	--	1	non
U.Texas	284	--	1	--	1	non
U.Oulu	125	1	16	1	1	non
XM2VTS	295	--	1	--	4	oui
Yale	15	1	3	6	1	non
Yale B	10	9	64	1	1	non

Tableau 3.1. Principales caractéristiques des bases de visages. Le tableau contient le nombre de personnes enregistré, le nombre de vues sous des poses et conditions d'illumination différentes, ainsi que le nombre de sessions au cours desquelles des vues d'une même personne ont pu être collectées. Les cas où l'un des éléments n'a pas été mesuré, ou était non contrôlé durant la prise de vue, est noté "--".

En parallèle, les protocoles d'évaluation XM2VTS ont essentiellement été introduits pour l'évaluation des algorithmes multimodaux (voix, visage) et 3D, et ceci pour l'authentification uniquement. Récemment, le NIST a publié les résultats du "Face Recognition Grand Challenge" [FRGC02] (Le grand défi de la reconnaissance faciale), un concours ouvert aux chercheurs qui se composait d'une double épreuve, le FRVT 2006 [Phi07] (Face Recognition Vendor Test, le test des fournisseurs de programmes de reconnaissance faciale) et le ICE (Iris Challenge Evaluation, le défi de l'iris).

Dans ce qui suit nous allons présenter en détails deux bases de données utilisées dans le cadre notre thèse, à savoir la base XM2VTS [Mes99] et la nouvelle base IV². Nous présenterons aussi le protocole d'évaluation de la phase 1 de développement de IV² [IV2].

3.2 La base XM2VTS [Mes99]

C'est une base multimodale développée au sein du projet européen ACTS, elle contient des images fixes, des séquences vidéo et des images faciales 3D de 295 personnes. Elle est utilisée pour la vérification d'identité. La base de données XM2VTS a été réalisée sur une longue période ce qui a permis d'obtenir plusieurs images d'une même personnes et donc une grande variabilité d'apparence (changement de coiffure, présence et/ou absence de lunettes, etc). Par contre, seules les expressions neutres ont été considérées.

La base de données XM2VTS contient 4 sessions. Une session "parole" où les personnes regardent vers le bas en lisant une phrase phonétiquement équilibrée. Une image de visage avec une bouche fermée a été choisie. Deux sessions avec et sans lunettes ont été acquises pour des sujets qui portent régulièrement des lunettes. La quatrième session correspond à une acquisition 3D de visages.

Pour la tâche de vérification, un protocole standard d'estimation de performances a été développé. Appelé «*Lausanne protocol splits randomly*», il consiste à séparer les individus en deux classes, client et imposteur. Le groupe client contient 200 sujets, alors que le groupe imposteur est divisé en 25 imposteurs pour l'évaluation et 70 imposteurs pour le test. Huit images des quatre sessions sont utilisées.

La base de données est divisée en trois ensembles : apprentissage, évaluation et test. L'ensemble apprentissage permet de construire les modèles de clients. L'ensemble test est

utilisé pour calculer les scores des clients et des imposteurs. En fonction de ces scores, un seuil est choisi afin de déterminer si une personne est acceptée ou non. D'après le protocole de *Lausanne* le seuil est choisi telle manière à ce que l'ensemble de test satisfasse certains niveaux de performances. Finalement l'ensemble test est sélectionné pour simuler un scénario d'authentification réel où l'identité de l'imposteur est inconnue au système. Enfin, l'ensemble évaluation est aussi utilisé pour les expériences de fusion pour l'apprentissage.

Nous avons utilisée la base XM2VTS pour évaluer notre approche d'extraction du visage et de ses caractéristiques.



Figure 3. 1. Exemple d'images extraites de la base XM2VTS.

3.3 La base IV² [IV2]

La base de données IV² a été développée dans le cadre du programme national TechnoVision, cofinancé par les ministères de la recherche et de la défense, lancé en 2004. Cette base contient des données intéressantes pour évaluer différents types d'algorithmes, notamment sur : la vérification par l'iris, par le visage 2D et 3D et par l'analyse de visages parlants, les approches multimodales telles que iris/visage ou voix/visage, etc. La campagne d'acquisition a commencé en août 2005. La base IV² comporte 482 identifiants, correspondant à 365 personnes, dont 268 personnes enregistrées sur une session, 77 enregistrées sur 2 sessions, et

20 enregistrées sur 3 sessions. Un ensemble de 52 sujets disjoint de la base précédente, a été défini pour la phase de développement. L'avantage de IV² par rapport aux autres bases est qu'elle contient plus d'informations sur chaque individu : 5 Expressions faciales, 5 poses, avec variation des conditions d'éclairage. D'autre part, chaque visage a été acquis avec plusieurs capteurs : un caméscope numérique de haute résolution pour l'acquisition du visage parlant, une Webcam pour une deuxième acquisition du visage parlant, 4 caméras matricielles CCD couleurs pour obtenir des informations de face et de profil permettre la reconstruction de ce visage en 3D, et enfin un scanner laser pour le visage 3D et caméra infrarouge pour les iris.

3.3.1 Protocoles d'acquisition IV²

Pour chaque session .

Scanner 3D

8 vues : 1 face neutre, puis un $\frac{3}{4}$ droit et un $\frac{3}{4}$ gauche neutre (face, gauche, droite), face yeux fermés, dans différentes expression (face souriante, face exprimant la surprise, face exprimant le dégoût et face neutre sans lumière continue du parapluie).

Webcam/Caméscope

La webcam et le caméscope démarrent en même temps. A ce moment-là le volontaire lit les différentes phrases affichées dans la cabine. A la fin de la lecture, la personne regarde à droite (profil gauche), à gauche (profil droit), au plafond et vers le plancher. Le caméscope reste allumé pendant le point n°3.

Bases stéréo

On commence en lumière maximum, la personne étant de face.

- 10 secondes en pleine lumière
- 10 à 12 secondes avec décroissance de la lumière (avec 3 ou 4 paliers). La personne étant toujours de face.
- 10 secondes maximum avec bascule sur les halogènes, 5 secondes le droit et 5 secondes le gauche, la personne étant toujours de face.
- 10 à 15 secondes finales, on remet l'éclairage total et on demande à la personne de réaliser un trajet visuel qui part du centre et qui va balayer les icônes présentes dans la

cabine. L'objectif étant de susciter des variations d'expressions (face souriante, face exprimant la surprise, face exprimant le dégoût) chez le volontaire.

Iris

2 sessions par personne pour chacun des yeux gauche et droit.

- La première session comprendra 10 images qualifiées de chaque oeil prises sans lunettes si le donateur porte des lunettes.
- La deuxième session sera prise avec lunettes, pour les donateurs en portant, et sans lunettes pour les autres, dans des conditions identiques à la première session.

Le tableau 3.1 donne quelques points de comparaison entre la base de données FRGC et celle développée dans le cadre du projet IV². Cette comparaison porte sur les aspects 3D. On peut constater au vu de ce tableau qu'une plus grande variabilité de données est proposée par la base IV². De plus, dans la mesure où nous souhaitons utiliser la complémentarité des informations 2D et 3D, il nous faut impérativement disposer de paires d'images acquises par un capteur stéréo. Ainsi, à partir des deux caméras nous pouvons avoir une information 2D (avec possibilité de redondance par utilisation des images gauches et droites) et en utilisant les deux images simultanément ainsi que les données de calibration nous pouvons remonter à l'information 3D pour tout ou partie de l'image.

Comparaison des données 3D	FRGC V1	IV ²
3D complet	non disponible	disponible
Multi-vues (poses)	(1 pose frontale)	(3 poses)
Illumination	contrôlée	contrôlée/incontrôlée
Expressions	2 expressions	5 expressions
2D et 3D	disponibles	disponibles
Image Stéréo	non disponible	disponible
Taille	4000 scans	5000 scans
Matériel (scanner 3D Minolta)	VI-910 (640×480)	VI-300 (400×400)

Tableau 3. 1 Comparaison des données 3D de la base FRGC V1 avec la base IV²

3.3.2 Evaluation préliminaire

3.3.2.1 Données de test

Les données utilisées dans cette phase de développement sont celles de 52 personnes. Ces données ont été acquises avant la mise au point du protocole d'acquisition final.

3.3.2.2 Extraction des données

Une campagne d'évaluations préliminaire a été réalisée à partir des données de la base de développement. L'obtention des visages 2D a nécessité une phase d'extraction à partir des vidéos disponibles (enregistrements caméscope). Dix images par personne ont été extraites (voir figure 3.2), de manière automatique à l'aide du logiciel **ImageGrab**. Le format d'image choisi est le format jpeg avec un facteur de qualité de 95. Les images sont espacées de 5 secondes à partir du début de la séquence.



Figure 3. 2. Exemple d'extraction de 10 images d'une personne dans la base de développement.

3.3.2.3 Annotation et normalisation

L'annotation manuelle des fichiers ainsi collectés (repérage des positions du nez et des yeux) a été réalisée sur plus de 500 images 2D. L'annotation constitue une phase préalable indispensable pour la normalisation des images. Les images sont ensuite normalisées à l'aide de l'algorithme de normalisation de la plateforme Biosecure 2005 [Bio04]. L'étape de normalisation consiste à réaliser les opérations suivantes:

- Transformation des images couleur en images à 256 niveaux de gris.

- rotations du visage dans l'image, de manière à ce que l'axe interoculaire soit horizontal,
- mise à l'échelle de l'image, de manière à ce que l'axe interoculaire soit égal à 70 pixels

Remarquons que la résolution finale des images a été fixée à 200×220 et qu'aucune égalisation ni normalisation de la luminance ne sont réalisées.



Figure 3.3. Normalisation des visages présentés (figure 3.1).

3.3.3 Protocoles

3.3.3.1 Comparaisons

Pour réaliser les comparaisons, il a été décidé de procéder plutôt par liste de comparaisons deux à deux à effectuer entre signatures. Ainsi, le logiciel de génération de la liste de comparaisons effectue un tirage aléatoire qui garantit un nombre de comparaisons inter et intra-classes identiques. 500 comparaisons dont 250 intra-classes ont été effectuées.

Le nombre de comparaisons sera choisi en fonction du score attendu : pour avoir un test statistiquement significatif, il faut que le nombre d'erreurs constatées soit au moins d'une dizaine. Ainsi, pour évaluer un EER de 0.05, il faut au moins 200 comparaisons intra-classe (et autant de comparaisons inter-classes).

3.3.3.2 Métriques de performances

Pour les tests de performance nous avons utilisé les notions suivantes :

Score : résultat du calcul de comparaison entre deux signatures biométriques.

Similarité : score normalisé entre 0 et 1, représentant le degré de ressemblance entre deux échantillons biométriques.

Distance : score représentant le degré de dissimilarité entre deux échantillons biométriques. Différentes distances sont classiquement utilisées par les algorithmes biométriques (L_1 , L_2 , cosinus, etc ...)

Normalisation du score : opération consistant à exprimer le score en termes de similarité. Différentes méthodes de normalisation peuvent être appliquées. Dans notre cas nous avons utilisé la normalisation min et max pour normaliser les scores, Nous notons le résultat absolu de la comparaison comme s , à partir d'un ensemble S de tous les scores pour la correspondance recherchée, et le score normalisé correspondant comme S_n .

Normalisation min-max : cette méthode transforme les scores issus de la mise en correspondance à l'intervalle $[0, 1]$. La quantité $\max(S)$ et $\min(S)$ spécifient les valeurs extrêmes de l'éventail des scores :

$$S_n = \frac{s - \min(S)}{\max(S) - \min(S)} \quad (3.1)$$

3.3.4 Conclusion

Dans ce chapitre, nous avons donné un aperçu global sur les bases de données existantes dans le domaine de la biométrie. Nous avons détaillé particulièrement les bases de données XM2VTS et IV², qui seront utilisées pour évaluer l'efficacité des algorithmes développés dans les chapitres suivants.

Chapitre 4

Extraction de visage et de ses caractéristiques

4.1 Introduction

Les performances des systèmes de reconnaissance de visage dépendent en grande partie de l'efficacité de la méthode de détection de visages utilisée. Dans le chapitre 2, nous avons mis en évidence l'intérêt d'une approche basée sur l'analyse de la couleur de la peau pour l'extraction de visages, bien que les méthodes permettant de le faire soient en général peu robustes aux variations d'éclairage comme nous l'avons signalé. D'autre part, le choix d'un espace optimal de représentation pour classifier les couleurs d'une image n'est pas évident. Parmi tous les espaces de couleur testés ces dix dernières années [Xu06], aucun n'émerge spécialement. Il semble donc intéressant d'étudier une approche différente. Dans ce chapitre, nous allons présenter une approche originale d'extraction d'un visage et de ses caractéristiques (yeux, nez, bouche) dans une image contenant un visage sur un fond uniforme. Cette approche est basée sur la segmentation de l'image en région et sur l'analyse de la couleur. L'algorithme de segmentation que nous avons développé possède les atouts suivants :

- Il permet d'obtenir de grandes régions de couleurs homogènes qui correspondent aux différentes régions du visage.

- Il est robuste aux variations d'illumination, grâce à notre choix d'utiliser l'espace couleur TLS, ainsi que des outils comme le gradient TLS.
- Il permet de classifier les pixels d'images en deux catégories : ceux ayant la couleur de la peau et les pixels ayant une couleur différentes. La discrimination entre ces deux catégories est rendue possible grâce à une segmentation morphologique de l'image couleur.

La robustesse de la détection de visage face aux variations de luminosité, de poses et d'expressions (surtout pour les caractéristiques du visage face aux expressions). Nécessité d'une bonne précision de la détection pour aborder la phase d'authentification dans les meilleures conditions.

Dans ce qui suit, nous décrivons notre approche pour segmenter une image couleur en régions et extraire les principaux éléments du visage. Cette approche est illustrée par de nombreux exemples.

4.2 Segmentation couleur du visage

La segmentation d'images fait partie des problèmes classiques de la vision par ordinateur. Une des applications principales de la segmentation est la reconnaissance de formes. En effet, la décomposition d'une image en éléments moins complexes grâce à une compréhension du contenu sémantique de la scène représentée permet d'identifier les objets qui s'y trouvent.

La segmentation consiste à réaliser une partition de l'image en sous-ensembles (régions) homogènes. Ainsi, une région correspond à un ensemble connexe de points de l'image ayant des propriétés communes (intensité lumineuse, texture, couleur...) qui les différencient des pixels des régions voisines. Cette partition est souvent difficile à réaliser car peu d'images sont constituées uniquement de zones homogènes. Il existe deux grandes classes de méthodes [Coc95], celles qui extraient les zones homogènes (régions) et celles qui extraient les zones de transition (contours).

La segmentation peut être utilisée aussi bien pour séparer les visages du fond que pour isoler certaines structures faciales. Plusieurs recherches ont été menées dans ce domaine sur des images en niveau de gris ou en couleur. La plupart des techniques de segmentation couleur de visages utilisent la couleur de peau pour réaliser la détection du visage. En effet,

la peau occupe une petite zone bien délimitée dans l'espace des couleurs, ce qui facilite la séparation entre les régions de peau et les autres régions de l'image. Nous avons vu dans le chapitre 2, que la distribution de la couleur de la peau est généralement représentée par un modèle paramétrique (tel que : le modèle gaussien simple, le modèle mélange gaussien, le modèle elliptique), ou bien par des règles explicites.

Notre approche de segmentation du visage est basée sur la méthode dite Ligne de Partage des Eaux (LPE) dans l'espace TLS. Ce choix se justifie par le fait que la LPE est très efficace pour segmenter des régions convexes et fermées, ce qui est le cas du visage humain. Tout d'abord nous donnons une présentation des outils morphologiques de segmentation en particulier la ligne de partage des eaux, qui nous seront utiles par la suite.

4.3 Segmentation par Ligne de Partage des Eaux

La Ligne de Partage des Eaux a été proposée initialement par Digabel et Lantuéjoul [Dig78], puis améliorée par Beucher et Lantuéjoul [Beu79]. Issue de la théorie de la morphologie mathématique, cette méthode est souvent utilisée en segmentation d'images.

Elle utilise le principe de l'immersion pour effectuer la segmentation de l'image en régions.

4.3.1 Principe de l'immersion

Soit I l'image sur laquelle on souhaite calculer la Ligne de Partage des Eaux. L'image I est considérée comme une surface topographique (figure 4.1). Imaginons que cette surface soit trouée aux emplacements des minima régionaux et qu'elle soit immergée dans un lac de manière à ce que le niveau des eaux monte à une vitesse constante. L'eau remplira progressivement les bassins du captage en commençant par le pixel ayant la plus basse altitude. Chaque fois que les eaux venant de deux minima différents entrent en contact, nous construisons un "barrage" pour les séparer. À la fin de la procédure d'immersion, chaque minimum sera complètement entouré par des barrages qui délimitent les eaux appartenant au bassin versant associé. Ces barrages constituent ce qu'on appelle la Ligne de Partage des Eaux.

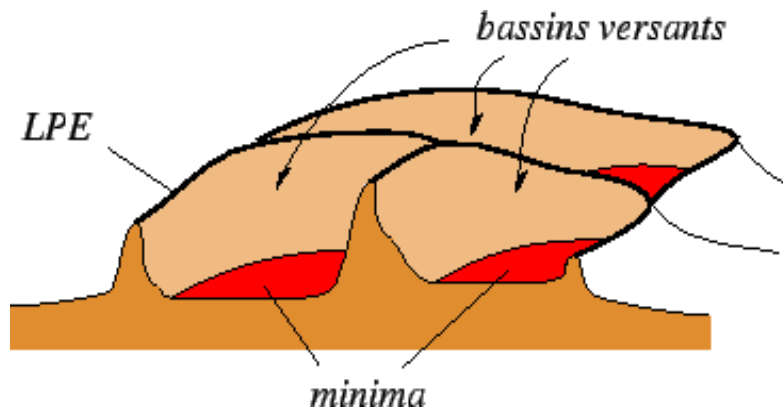


Figure 4. 1. Minima, bassins versants et ligne de partage des eaux.

4.3.1.1 Construction de la LPE

La formulation de la ligne de partage des eaux en termes d'immersion présente l'avantage de fournir un algorithme direct pour sa construction. Avant de formaliser l'immersion, donnons quelques définitions [Beu79] :

Définition 1 : Une image de niveau de gris est une application $f: D \rightarrow N$, où D est l'espace de points ($D \subseteq Z^2$). Soit p un pixel, $f(p)$ est la valeur du niveau de gris de p .

Définition 2 [Chemin géodésique] : considérons l'ensemble X de la figure 4.2 Soient deux points s, t appartenant à cet ensemble. Une ligne ou un arc d'extrémités s et t sera appelé **chemin géodésique** s'il vérifie les conditions suivantes :

- être totalement inclus dans l'ensemble X ;
- correspondre au plus petit parcours entre s et t .

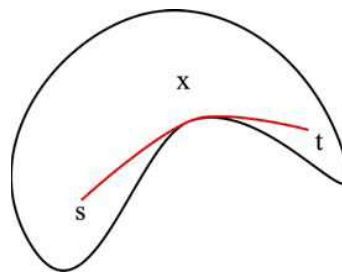


Figure 4. 2. Le plus court chemin géodésique entre s et t .

Définition 3 [Zone d'influence géodésique - ZI] : Soit, dans R , l'ensemble $X = \{X_i, i \in I\}$ formé de I composantes connexes et compactes, toutes incluses dans le compact Y . La zone

d'influence géodésique de X_i dans Y se définit comme le lieu des points de Y qui sont géodésiquement plus proches de X_i que toutes les composantes connexes de X .

$$ZI_Y(X_i) = \{p \in Y / d_Y(p, X_i) < d_Y(p, X_j) \forall j \neq i\} \quad (4.1)$$

Où d_Y est la distance géodésique dans l'ensemble Y (figure 4.3).

Définition 4 [Squelette par zone d'influence - SKIZ] : C'est la frontière entre les différentes zones d'influence (voir figure 4.3.).

$$SKIZ_Y(X_1, \dots, X_n) = Y \setminus \bigcup_{i=1}^n ZI_Y(X_i) \quad (4.2)$$

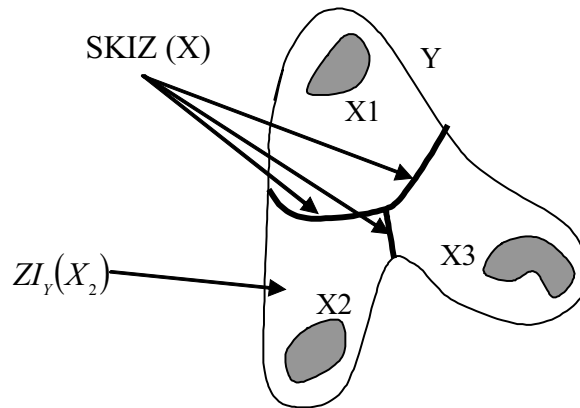


Figure 4. 3. Exemple de SKIZ géodésique.

Pour formaliser l'immersion, nous utilisons les notations suivantes :

- $f_h \in [h_{min}, h_{max}]$ est l'ensemble des pixels $p \in D$, tel que $f(p) \leq h$.
- h_{min} et h_{max} représentent le minimum et le maximum de f , respectivement.
- $CB(M)$ est le bassin versant associé au minimum local M .
- Les points du bassin versant dont l'altitude est inférieure ou égale à h sont définis par :

$$CB_h(M) = CB(M) \cap f_h \quad (4.3)$$

- $X_h = \bigcup_i CB_h(M_i)$ représente l'union de tous les bassins versants qui ont une altitude inférieure ou égale à h .
- $MINR_h(f)$ est l'ensemble union des minima régionaux d'élévation h .

La simulation de l'inondation (immersion) se fait par construction des bassins, étape par étape suite à une augmentation progressive du niveau de l'eau.

La première inondation a lieu lorsque l'eau atteint le niveau $h_{min} : X_{h_{min}} = f_{h_{min}} = MINR_{h_{min}}(f)$. Soit Y une composante connexe de f_{h+1} . Lorsque le niveau d'eau monte du niveau h au niveau $h+1$, il y a trois cas de figure, pour gérer cette montée : f_{h+1} .

- $Y \cap X_h = \emptyset$; dans ce cas Y est un nouveau minimum régional d'altitude $h+1$.
- $Y \cap X_h \neq \emptyset$ et est connexe ; dans ce cas Y est la dilatation du bassin versant $CB_h(Y \cap X_h)$. Ainsi $CB_{h+1}(Y \cap X_h) = ZI_Y(Y \cap X_h)$.
- $Y \cap X_h \neq \emptyset$ et n'est pas connexe (on note Z_i ses composantes connexes). Dans ce cas de figure Y est la réunion des eaux provenant de plusieurs minima régionaux. Comme cette jonction n'est pas autorisée ; il faut donc construire la ligne de partage des eaux, séparant ces différents bassins. Pour cela on construit des zones d'influence géodésique :

$$X_{h+1} = MINR_{h+1}(f) \cup ZI_{h+1}(X_h) \tag{4.4}$$

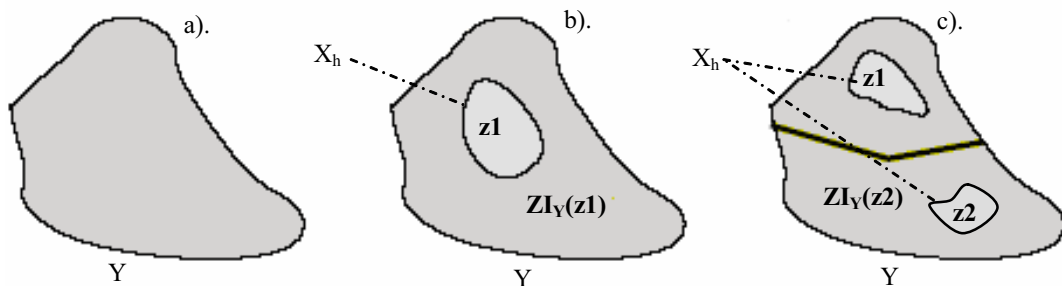


Figure 4. 4. Les différents cas de l'immersion.

4.3.2 Ligne de Partage des Eaux (d'après Vincent et Soille)

La première implémentation de l'algorithme LPE par immersion a été introduite par Vincent et Soille en 1991 [Vin91]. L'algorithme proposé met en application le principe de l'immersion. L'idée est de considérer des seuils successifs de l'image et de calculer des zones d'influence géodésiques d'un niveau à l'autre le plus rapidement et le plus exactement possible. Pour avoir un accès direct aux pixels d'un niveau de seuil donné, un tri initial des pixels dans l'ordre des niveaux de gris croissants est effectué. Cette première étape exploite le fait que le nombre de niveaux de gris possible dans l'image est fixe : le tri est ainsi effectué en temps linéaire par rapport au nombre de pixels. La deuxième étape, l'inondation, met en jeu des calculs successifs de zones d'influence géodésiques. Nous allons détailler cet algorithme dans la partie implémentation.

L'algorithme de LPE génère le plus souvent une sur-segmentation importante. La sur-segmentation se produit à cause du grand nombre de minima locaux non significatifs. Il suffit de petites fluctuations du relief pour produire autant de minima locaux dans l'image du gradient. Afin de limiter la sur-segmentation, une solution consiste à utiliser des marqueurs. L'inondation débute alors non par les minima de l'image, mais par l'ensemble des marqueurs choisis, un par région à segmenter. L'inconvénient de cette méthode réside dans la difficulté d'initialiser les marqueurs. Ceci a conduit à l'apparition des approches de Ligne de Partage des Eaux par fusion de régions.

Pour contrer la sur-segmentation, plusieurs versions modifiées de l'algorithme originel de la LPE ont été proposées. Citons comme exemple l'algorithme de la LPE modifié par Andrade [And97]. En effet, Andrade s'affranchit de l'initialisation des minima régionaux car ceux-ci sont construits au fur et à mesure de l'inondation. Le problème de la sur-segmentation est réglé au moment de la jonction de deux régions : c'est-à-dire au moment où l'on érige un barrage qui sera la frontière entre deux régions. La croissance de la région continuera, mais on décidera au moment de cette jonction si on fusionne ou non les deux régions en une seule. C'est cette approche que nous avons retenue. Nous donnerons plus de précisions sur le processus de fusion dans la partie implémentation de l'algorithme.

4.4 Approche segmentation de couleur proposée

Un des objectifs que nous nous sommes fixés est de mettre en place des outils efficaces pour la segmentation de visage. Ces outils faciliteront l'interprétation de l'image pour l'extraction de visage et des régions caractéristiques. Pour atteindre cet objectif, il faut tout d'abord choisir un espace couleur adéquat, ainsi que l'opérateur gradient le mieux adapté aux besoins spécifiés. Nous avons choisi d'utiliser l'espace couleur TLS (Teinte, Luminance, Saturation) ainsi que gradient TLS pour avoir une robustesse aux variations d'illumination.

Nous proposons de modifier la méthode de fusion des bassins versants obtenus à partir de l'algorithme de lignes de partage des eaux (LPE) en ajoutant un critère basé sur la teinte. Notre algorithme de segmentation de visage s'exécute sur une image de la norme du gradient et effectue une croissance de régions en simulant l'inondation de l'image de la même manière que l'algorithme de la Ligne de Partage des Eaux. Dans ce cas, chaque minimum local de la norme du gradient est considéré comme un germe de région, il représente le cœur d'une région. Au fur et à mesure de l'inondation, les pixels sont agrégés aux supports des régions.

Les différentes étapes qui constituent notre algorithme peuvent être résumées comme suit :

- Transformation de l'image de l'espace RVB à l'espace TLS
- Détermination de l'image de la norme du gradient TLS
- Simulation de l'inondation
- Fusion des bassins versants.

Dans les sections suivantes, nous allons étudier en détail chacune de ces étapes, nous présenterons les algorithmes associés, ainsi que les résultats obtenus sur des images réelles.

4.4.1 Représentation de l'espace TLS

Le système TLS (Teinte, Luminance, Saturation) est proche de la perception humaine des couleurs. L'intérêt essentiel de cette représentation est qu'elle assure une décorrélation entre l'information de couleur et les autres composantes. Au lieu de donner les 3 composantes rouge, vert, bleu, l'utilisateur choisit une teinte (couleur dans un cercle

chromatique allant du rouge au violet, en passant successivement par l'orange, le jaune, le vert et le bleu), une luminosité (couleur plus ou moins claire) et une saturation (couleur plus ou moins intense). Dans un premier temps, on détermine la couleur souhaitée et la luminosité définit la quantité de noir ou de blanc dans la couleur sélectionnée. Cela permet de distinguer les couleurs claires des couleurs sombres. La saturation permet de mesurer la pureté des couleurs. Par exemple un rouge et un rose ne se distinguent que par la composante saturation, alors que dans l'espace RVB, il y a une corrélation élevée entre les trois composantes.

On calcule d'abord trois composantes YC_1C_2 , à l'aide de la transformation matricielle :

$$\begin{bmatrix} Y \\ C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1 & -1/2 & -1/2 \\ 0 & \sqrt{3}/2 & -\sqrt{3}/2 \end{bmatrix} \begin{bmatrix} R \\ V \\ B \end{bmatrix} \quad (4.5)$$

où Y est l'intensité et (C_1, C_2) sont des composantes de chrominance.

Puis, les composantes Teinte (T), Luminance (L) et Saturation (S) sont ensuite obtenues par les équations suivantes :

$$L = Y = (R + V + B) / 3 \quad (4.6)$$

$$S = \sqrt{C_1^2 + C_2^2} \quad (4.7)$$

$$C_3 = \text{atan}(C_1/C_2) \quad (4.8)$$

Si $(C_3 > \pi/2)$

$$T = ((\pi/2 - C_3) + 2 * \pi) \quad (4.9)$$

Sinon

$$T = (\pi/2 - C_3) \quad (4.10)$$

On note que la Teinte prend des valeurs cycliques qui peuvent être représentées sur un cercle (cf. figure 4.5):

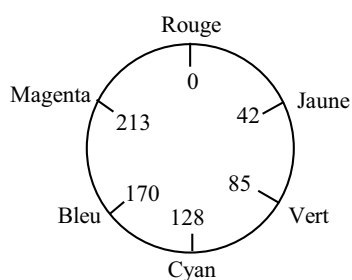


Figure 4. 5. Représentation numérique circulaire de la teinte

4.4.2 Détermination de l'image de la norme du gradient TLS

Le calcul du gradient d'une image couleur est une étape de prétraitement dont l'objectif est de faciliter les étapes ultérieures de segmentation. Comme pour les images en niveaux de gris, nous cherchons à mettre en évidence les variations locales attachées aux pixels. De nombreuses méthodes ont été proposées dans la littérature [Coc95]. Nous nous limiterons à la présentation de l'approche vectorielle de Di Zenzo [DiZ86]. C'est celle que nous emploierons dans notre implémentation.

4.4.2.1 Approche par gradient vectoriel de Di Zenzo

L'approche proposée par Di Zenzo est, le plus souvent, utilisée dans les méthodes demandant la détermination de contours multi-composantes. Elle relève d'une stratégie essentiellement vectorielle qui traite en une seule opération l'ensemble des composantes. Le pixel est alors considéré comme un vecteur (à trois composantes dans le cas des images couleur).

Le gradient scalaire est souvent mesuré par la variation d'intensité le long de la ligne de plus grande pente. Di Zenzo reprend cette notion pour définir le gradient vectoriel. Sa stratégie consiste à considérer deux pixels voisins (A , A_v), dans l'espace de l'image, à une distance **unitaire** l'un de l'autre. Leur position relative est repérée par l'angle θ (figure 4.6-a).

Il estime alors, en fonction de cet angle θ , la distance séparant ces deux vecteurs dans l'espace des composantes. La figure (3.6-b) illustre cette modélisation dans l'espace TLS.

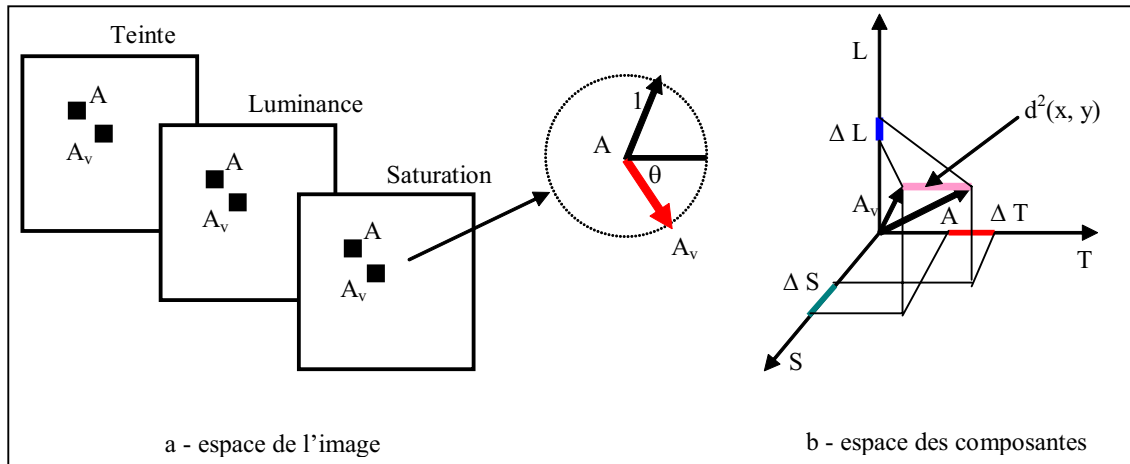


Figure 4. 6. Modélisation du gradient multi composantes avec l'approche Di Zenzo.

La stratégie de Di Zenzo consiste à calculer l'angle θ qui maximise la distance $d^2(x, y)$.

La distance $d^2(x, y)$ s'exprime alors par :

$$d^2(x, y) = p \cos^2 \theta + q \sin^2 \theta + 2t \sin \theta \cos \theta \quad (4.11)$$

avec :

$$p = \left(\frac{\partial T}{\partial x} \right)^2 + \left(\frac{\partial L}{\partial x} \right)^2 + \left(\frac{\partial S}{\partial x} \right)^2 \quad q = \left(\frac{\partial T}{\partial y} \right)^2 + \left(\frac{\partial L}{\partial y} \right)^2 + \left(\frac{\partial S}{\partial y} \right)^2 \quad (4.12)$$

$$t = \frac{\partial T}{\partial x} \cdot \frac{\partial T}{\partial y} + \frac{\partial L}{\partial x} \cdot \frac{\partial L}{\partial y} + \frac{\partial S}{\partial x} \cdot \frac{\partial S}{\partial y} \quad (4.13)$$

Le maximum de la distance est obtenu pour (θ_G) qui est la direction du gradient et le module carré du gradient ($|G_{vect}|$) :

$$\theta_G = \frac{1}{2} \arctan \left(\frac{2t}{p-q} \right) \quad (4.14)$$

$$|G_{vect}|^2 = \frac{1}{2} \left\{ p + q + \sqrt{(p+q)^2 - 4(pq - t^2)} \right\} \quad (4.15)$$

Mise en œuvre

Le gradient de Di Zenzo repose sur le calcul des quantités p , q et t qui font intervenir les dérivées selon x et y dans chacun des plans. Ces quantités marginales peuvent donc être estimées en utilisant le gradient vectoriel TLS qui est défini par les trois gradients marginaux des composantes teinte, luminance et saturation. Pour calculer ces derniers, nous avons utilisé la méthode de Carron [Car95]. En effet, notre étude de la littérature scientifique sur les techniques de segmentation dans l'espace TLS nous a permis de constater l'intérêt des travaux de Carron pour la segmentation dans cet espace, notamment en présence de bruit : « l'information de Teinte possède une fiabilité variable qui dépend du niveau de corrélation inter-composantes du bruit et, surtout, du niveau de Saturation [Car95] »

On peut résumer ceci en définissant trois concepts sur la Saturation permettant de caractériser la pertinence de la Teinte :

- Saturation “faible” : Teinte non significative.
- Saturation “moyenne” : Teinte peu significative.
- Saturation “forte” : Teinte significative.

L'originalité de cette méthode réside dans le fait que le calcul du gradient est fonction de la pertinence de la teinte. En effet, la teinte est une information dont la pertinence est variable : lorsque la couleur est fortement saturée, la teinte est une information fiable, peu sensible au bruit, et, à l'inverse, lorsque la saturation est faible, la teinte est très sensible au bruit et donc sa pertinence est mauvaise. En conclusion, privilégier la teinte lorsqu'elle est pertinente, permet de ne pas prendre en compte les zones d'ombre.

Un coefficient $\alpha(S)$, fonction de la saturation, est donc construit, pour mesurer la pertinence de la Teinte. Si la Saturation n'intervient pas dans le calcul du gradient de la Teinte, il n'y aura pas de contour entre deux zones ayant des teintes identiques, pertinentes et de Saturations différentes (exemple une zone rouge et une zone rose). Cette fonction est définie algébriquement par la relation suivante :

$$\alpha(S) = 1/\pi[\pi/2 + \arctg(\beta(S - S_0))] \quad (4.16)$$

Dans laquelle :

- S_0 permet de définir un niveau de pertinence moyen de la Teinte lié à un niveau de Saturation.
- β règle la pente de la fonction autour du niveau S_0 et permet de doser le mélange dans les situations où la pertinence accordée à l'information de Teinte est moyenne (figure 4.7).

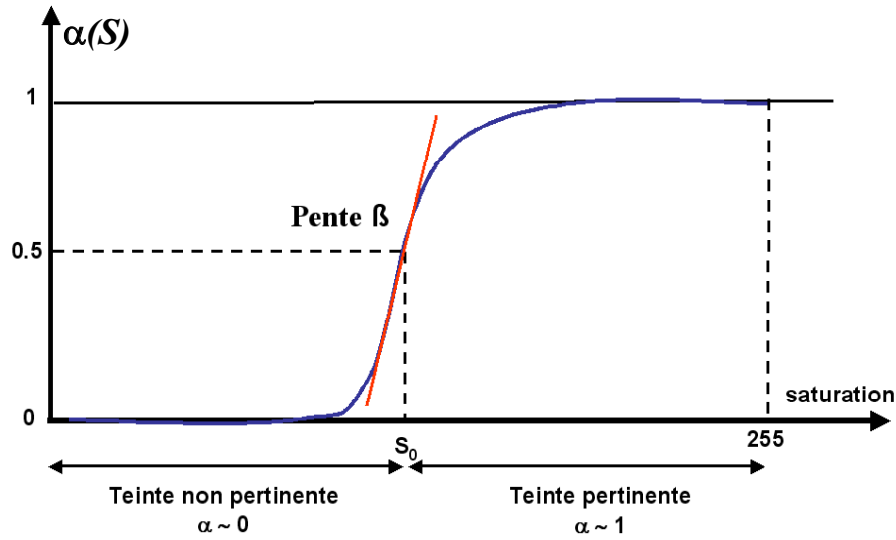


Figure 4. 7. Mesure de la pertinence de la teinte

Le coefficient pondérateur $\alpha(S)$ doit être une mesure globale des pertinences des Teintes de tous les pixels intervenants dans le calcul des gradients des trois composantes : Teinte, Saturation et Luminance. Le calcul du gradient en x et y est alors effectué de la manière suivante :

$$\begin{aligned}
 G_x^T &= \Delta T(T_i, T_j) \cdot M_{Sobel}^x, & G_y^T &= \Delta T(T_i, T_j) \cdot M_{Sobel}^y \\
 G_x^L &= (1 - p(S_i, S_j)) \cdot (L_i, L_j) \cdot M_{Sobel}^x, & G_y^L &= (1 - p(S_i, S_j)) \cdot (L_i, L_j) \cdot M_{Sobel}^y \\
 G_x^S &= (1 - p(S_i, S_j)) \cdot (S_i, S_j) \cdot M_{Sobel}^x, & G_y^S &= (1 - p(S_i, S_j)) \cdot (S_i, S_j) \cdot M_{Sobel}^y
 \end{aligned} \quad (4.17)$$

avec

M_{Sobel}^x, M_{Sobel}^y : l'opérateur gradient de Sobel en x et y

$$\Delta T(T_i, T_j) = p(S_i, S_j) \cdot |T_i - T_j|_{\text{modulo } 255} \quad (4.18)$$

$p(S_i, S_j) = \sqrt{\alpha(S_i) \cdot \alpha(S_j)}$ est la moyenne géométrique des coefficients pondérateurs entre deux pixels.

$|T_i - T_j|_{255}$ est la mesure de l'écart de Teinte entre les deux cœurs de régions (calculée modulo 255, afin de résoudre le problème de la représentation circulaire).

Les figures ci-dessous montrent un exemple des trois images T, L, S pour deux images avec des éclairages et des fonds différents. La figure (4.8b) illustre la capacité de la composante Teinte à éliminer les contours dus aux variations d'éclairage (flèche 1) et à ne tenir compte que des transitions significatives entre le fond et le visage (flèche 2). Dans ce cas, le facteur de pondération $\alpha(S)$ va automatiquement privilégier le gradient de Teinte par rapport aux autres composantes. Par contre dans la figure (4.9b), on voit bien que l'information de Teinte n'est pas pertinente (les transitions visage-fond sont invisibles), et dans ce cas, le facteur de pondération va plutôt privilégier la composante de Luminance.

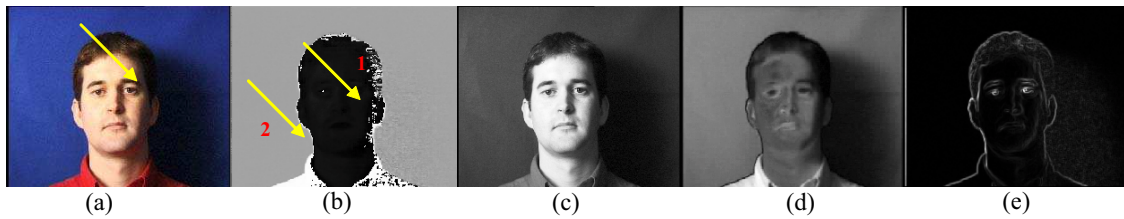


Figure 4. 8. Les différentes composantes : teinte (b), luminance (c), saturation (d), gradient(e).

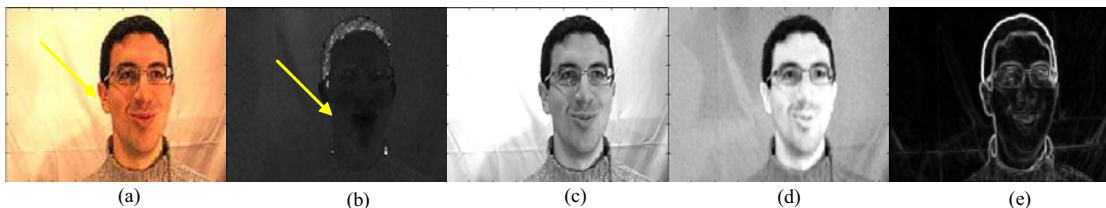


Figure 4. 9. Les différentes composantes : teinte (b), luminance (c), saturation (d), gradient(e).

4.4.3 Simulation de l'inondation d'après Vincent

4.4.3.1 Description générale

Nous avons utilisé l'algorithme de LPE proposé par [Vin91] pour créer les supports des régions. Cet algorithme est appliqué sur l'image de la norme du gradient couleur obtenue dans l'étape précédente, il est décomposé en deux étapes :

a) Etape de tri

Elle consiste à trier les pixels de l'image par ordre croissant de leur gradient. Parmi les méthodes de tri existantes, Vincent et Soille ont choisi l'algorithme proposé par E.J. Isaac et R.C Singleton [Isa56]. Il s'agit d'un algorithme de tri distributif qui fait appel à des calculs d'adresses. Cette technique suppose que les données à trier soient des entiers. Il est donc nécessaire de normaliser le gradient avant le tri. On choisit ici $[0,255]$ comme intervalle de normalisation.

L'algorithme de tri peut être décomposé en trois étapes :

- Détermination du tableau de distribution en fréquence des normes du gradient dans I , noté H_I l'histogramme de I :

$$\forall h, H_I(h) = \text{nombre de pixels de } I \text{ ayant pour gradient la valeur } h$$

- Détermination de l'histogramme cumulé HC_I de I :

$$\forall h, HC_I(h) = \text{nombre de pixels de } I \text{ ayant pour gradient la valeur } \leq h$$

- Parcours exhaustif de l'image I , au cours duquel tout pixel est mis directement dans une cellule dans le tableau trié. Cette cellule est repérée grâce à l'histogramme cumulé.

Notons alors n le nombre de pixels de l'image I , et h_{min} et h_{max} la plus petite et la plus grande valeur de la norme du gradient respectivement. Le couple constitué du tableau trié de pointeurs de pixels (I_I) et de l'histogramme cumulé permet, au cours de la phase d'immersion, un accès direct aux pixels d'une norme de gradient donnée.

b) Etape d'immersion

Une fois le tri effectué, on passe à l'inondation progressive des bassins versants de l'image. Supposons que cette inondation ait été accomplie jusqu'à un niveau donné h . À chaque niveau h la reconstruction géodésique est réalisée grâce à une file d'attente. Lorsque l'inondation a atteint le niveau h , tous les bassins versants déjà découverts ont un label grâce au classement dans un tableau. Les pixels du niveau $h+1$ sont obtenus directement à partir de ce même tableau. L'algorithme leur affecte une valeur que l'on notera MASK. Parmi ces pixels à la valeur MASK, ceux qui ont un pixel déjà étiqueté dans leur

voisinage (8 connexités) sont placés dans la file d'attente. À partir de ceux-ci, on peut donc se propager en largeur d'abord dans les régions à la valeur MASK. Ainsi les zones d'influence sont construites une fois que la file d'attente devient vide. Par ailleurs, il est nécessaire d'effectuer un deuxième passage sur l'image pour affecter un label aux pixels de niveau $h+1$ sans labels provenant des minima $h+1$. En effet, ces minima sont disjoints des bassins versants de niveau h . Un label particulier WSHED est alors affecté aux pixels où les bassins versants essaient de fusionner. Il désigne les pixels où se trouve la ligne de partage des eaux.

Quelques exemples (figure 4.10-a) représentent des visages en couleur auxquels nous avons appliqué l'algorithme détaillée de notre l'implémentation de l'algorithme est donné en annexe A) et où nous avons initialisé l'algorithme avec tous les minima locaux de la norme du gradient TLS.

Les bassins versants obtenus associés sont représentés dans les figures (4.10-b) respectivement. On note tout d'abord que l'algorithme LPE permet d'obtenir des contours fermés, ce qui fait sa grande force. Par contres les régions sont sur-segmentées et ne peuvent pas être utilisées telles quelles pour analyser le visage.

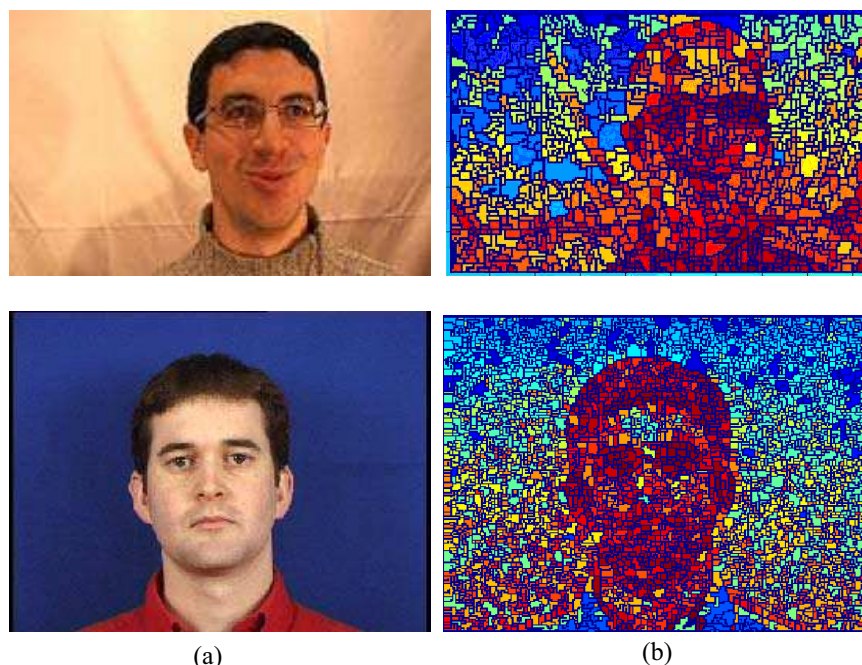


Figure 4. 10. Images originales (a) et résultats de l'immersion (b).

4.4.4 Fusion des bassins versants

4.4.4.1 Principe

Pour supprimer les minima locaux non significatifs nous avons employé la méthode de fusion des bassins versants proposée par Andrade [And97] qui n'est autre qu'une version modifiée de l'algorithme LPE non hiérarchique de Vincent. En effet, Andrade introduit dans l'algorithme originel de Vincent deux nouveaux paramètres pour contrôler la fusion des bassins versants : l'aire (nombre de pixels dans un bassin) et la profondeur (norme du gradient du cœur du bassin). Ces deux paramètres sont combinés pour définir une règle de fusion qui permet d'empêcher ou d'autoriser la fusion des bassins versants. Cette règle de fusion est ajoutée dans l'algorithme LPE à l'endroit où deux bassins entrent en contact, comme expliqué ci-dessous :

Au niveau h , les supports des régions existantes sont étendus, par dilatation conditionnelle des supports des régions au niveau $h-1$, conditionnellement au niveau h . À chaque dilatation de niveau h , chaque pixel ayant au moins un voisin déjà étiqueté est traité. Les étiquettes d'un pixel dépendent des étiquettes de leurs voisins. Si tous les voisins ont la même étiquette, le pixel prendra cette étiquette. S'il y a au moins deux étiquettes différentes, le pixel prendra l'une des étiquettes et les deux régions seront candidates à la fusion. L'aire et la profondeur des bassins associés à ces deux régions sont examinées : si pour l'un des bassins, l'un ou l'autre de ces attributs est inférieur à des seuils déterminés, le bassin le plus profond absorbe l'autre. Cette méthode suppose que les pixels du niveau courant aient été intégrés au support de la région, c'est-à-dire que toutes les dilatations des bassins du niveau courant aient été effectuées, afin que l'aire totale soit prise en compte dans la fusion. Par ailleurs, un bassin absorbé n'absorbe jamais un autre bassin.

La figure 4.11 représente un exemple typique illustrant le principe de la fusion basée sur les paramètres d'aire et de profondeur des bassins versants expliquée ci-dessus. Elle montre un profil unidimensionnel, partiellement inondé, à un niveau arbitraire. Imaginons que le relief soit lentement inondé à partir de ses minima locaux. Après un certain temps, le niveau d'inondation atteint la ligne de séparation sur le flanc droit du bassin versant marqué par son minimum M . À partir de ce moment, l'eau se déverse dans un autre bassin, la profondeur de ce bassin est donnée par la différence entre le niveau d'inondation et la hauteur de son minimum M .

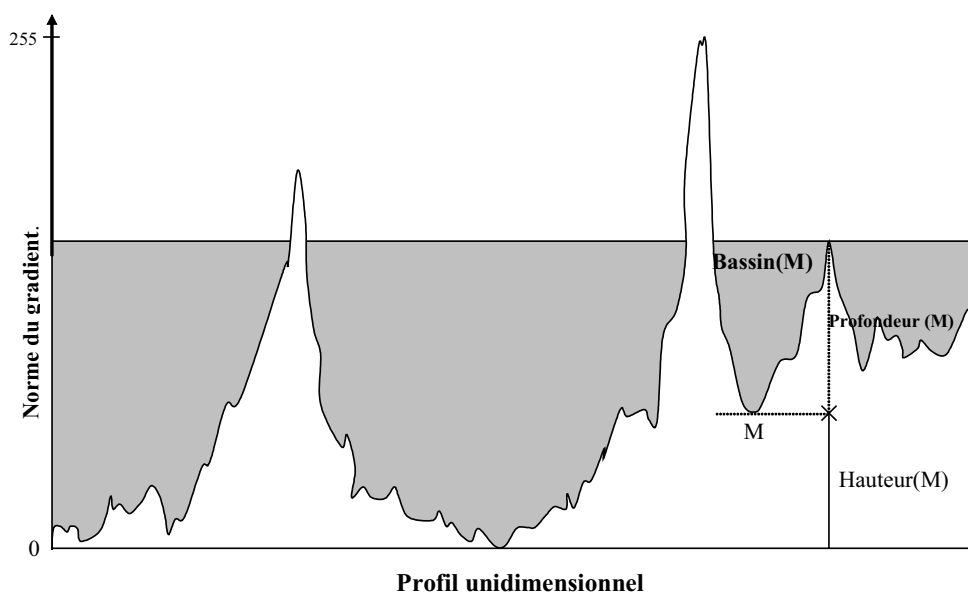


Figure 4. 11. Principe de la fusion des bassins versants.

4.4.4.2 Notre approche

Nous proposons de modifier la règle de fusion d'Andrade en supprimant le critère de l'aire des régions et en le remplaçant par le critère de la Teinte. En effet, dans notre application l'aire n'est pas un critère pertinent car les tailles des visages sont différentes d'une image à une autre, par contre la teinte constitue un descripteur discriminant de la couleur de la peau d'où l'intérêt de son utilisation. Ainsi, le critère de fusion que nous proposons combine la pertinence de la teinte (critère de Carron) avec la profondeur (norme du gradient du cœur du bassin). Nous avons donc utilisé la différence de niveaux entre les cœurs des bassins candidats à la fusion au lieu d'utiliser les informations couleur liées aux pixels. L'algorithme de fusion proposé est détaillé en annexe A.

L'idée de base est de former un seul critère scalaire (noté C) d'homogénéité mixant de manière pondérée les informations apportées par la Teinte et la Luminance. Or, comme nous l'avons vu précédemment (§ 4.4.2.2), la fiabilité de la Teinte est liée à l'information de Saturation S . Donc, l'utilisation du coefficient $\alpha(s)$ va permettre de pondérer les informations de Teinte et de Luminance (notées respectivement T et L) dans le critère d'homogénéité C .

Ce critère va indiquer la (ou les) composante(s) qui vont être utilisées pour la segmentation en régions. La figure 4.12 résume la stratégie de choix de ces composantes suivant la valeur du coefficient α et donc suivant la Saturation.

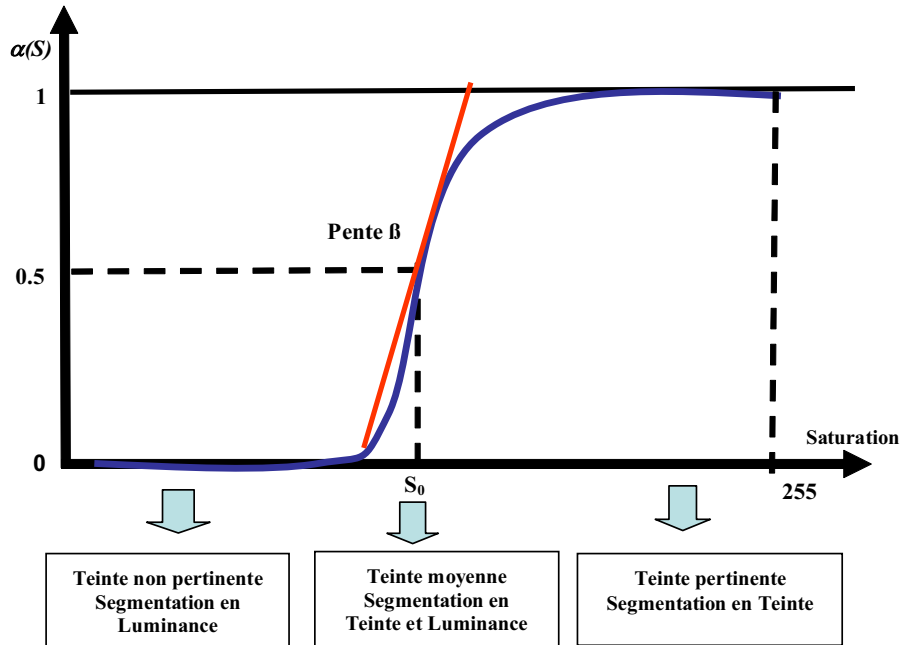


Figure 4. 12. Stratégie d'utilisation des composantes TLS.

Soient S_n , T_n et L_n les informations de Saturation, Teinte et Luminance correspondant au cœur du n -ième bassin et B_1 et B_2 deux bassins candidats à la fusion. Notre critère de fusion est défini par :

$$C(B_1, B_2) = p(S_1, S_2) \cdot \left| \left[T_1 - T_2 \right]_{255} \right| - (1 - p(S_1, S_2)) \cdot |L_1 - L_2| \leq \text{Seuil}_C \quad (4.19)$$

Où :

- $p(S_1, S_2) = \sqrt{\alpha(S_1) \cdot \alpha(S_2)}$ est la moyenne géométrique des coefficients pondérateurs entre deux cœurs de régions.
- Seuil_C est le seuil d'agrégation déterminant le prédicat d'homogénéité.

Le seuil d'agrégation Seuil_C doit être adaptatif pour prendre en compte le problème de la pertinence de la Teinte. Il est intéressant de pouvoir disposer de différentes valeurs de seuil d'agrégation suivant que la méthode de segmentation utilise la Luminance ou la Teinte, en d'autres termes, un seuil dynamique qui varie en fonction de la pertinence de la teinte.

Le seuil composite de fusion (noté $Seuil_c$) pour les deux bassins B_1 et B_2 est défini de la manière suivante [Car95] :

$$Seuil_c(B_1, B_2) = p(S_1, S_2) \cdot S_T + (1 - p(S_1, S_2)) \cdot S_L \quad (4.20)$$

Où S_L et S_T sont deux seuils déterminés après analyse de l'histogramme de l'histogramme de composante Saturation.

Pour de faibles niveaux de Saturation (§ 4.4.4.3), le critère d'homogénéité étant alors essentiellement basé sur la Luminance, le seuil composite d'agrégation va prendre la valeur S_L . A l'inverse, pour de fortes Saturations, le critère d'homogénéité étant formé par la composante Teinte, le niveau du seuil d'agrégation correspond à la cette composante (c-à-d. S_T).

L'algorithme de fusion de deux bassins B_1 et B_2 est résumé comme suit : (S_{prof} est le seuil qui détermine la profondeur des bassins) :

```

h1 = M - B1.profondeur
h2 = M - B2.profondeur
Si (C(B1 , B) <= Seuil_c(B1 , B)) &&( h1 < S_prof || h2 < S_prof)
    { Si (haut1 >= haut2)
        //le bassin B1 absorbe le bassin B2

            Mettre à jour le bassin B1
    Sinon (haut1 < haut2)
        Le bassin B2 absorbe la bassin B1

        Mettre à jour le bassin B2 }

```

La figure 4.13 illustre sur deux exemples la segmentation obtenue par à l'aide de notre méthode. Nous remarquons que le visage est bien segmenté en régions fermées et distinctes correspondant aux différentes parties du visage (yeux, nez, bouche et sourcils) cette segmentation est obtenue manuellement.

	S _{prof}	S _L	S ₀	S _T
Figure 4.13 (1)	25	10	50	150
Figure 4.13 (2)	15	50	160	176

Tableau 4. 1. Les différents seuils.

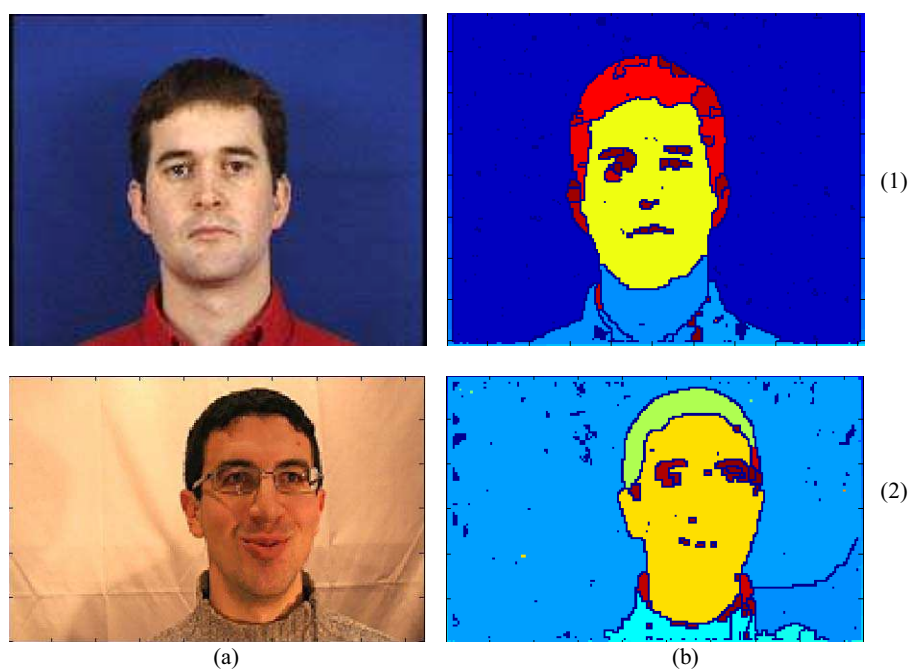


Figure 4. 13. Exemples de segmentation de visage : (a) images originales (b) images LPE modifié.

4.4.4.3 Détermination des paramètres S_0 et S_L et S_T

Pour permettre de déterminer les valeurs des paramètres S_0 , S_L et S_T , il est nécessaire de faire une étude préalable de l'histogramme de la Saturation. Le problème principal de toute méthode basée sur le seuillage de l'histogramme est le choix des seuils (ou de l'intervalle de seuillage). Avec un intervalle trop large, on obtient des *faux positifs*, c.-à-d. l'image seuillée contient des pixels qui ne font pas partie des objets d'intérêt ; généralement il s'agit de bruit, ou des structures d'une autre nature, qui ont un niveau de gris proche de celui des objets recherchés. Avec un intervalle trop étroit, on obtient des *faux négatifs*, c.-à-d. certains objets d'intérêt n'apparaissent pas, ou que partiellement, dans l'image seuillée.

Il existe de très nombreuses méthodes de seuillage d'un histogramme [Ots79][Tsa95] [Fu81]. La plupart de ces méthodes s'appliquent correctement si l'histogramme contient réellement des pics séparés. De plus, ces méthodes ont très souvent été développées pour traiter le cas particulier de la segmentation en deux classes et leur généralité face aux cas multi-classes n'est que très rarement garantie.

Algorithme :

Pour régler ces seuils de manière automatique, nous avons appliqué la technique de multi seuillage de Cheng et al. [Che97] sur l'histogramme de saturation. À l'origine, cette méthode permet d'obtenir un nombre désiré de classes à partir de l'histogramme de niveau de gris de l'image. Son principe est le suivant : l'histogramme original est lissé (figure 3.14-a) par une fonction gaussienne ayant un paramètre de diffusion ($\sigma = 1$) de telle sorte que les pics principaux de l'histogramme soient visibles. Le nombre de pics considérés dans l'histogramme lissé doit être égal au nombre de sommets dans l'histogramme original. Ensuite un filtrage passe-bas est appliqué sur l'histogramme lissé afin d'éliminer les sommets non significatifs, jusqu'à ce que le nombre de sommets restants soit égal au nombre désiré de classes. Si par contre, le nombre de sommets de l'histogramme lissé est inférieur au nombre de classes, alors un filtrage passe-haut est appliqué pour produire plus de sommets dans l'histogramme, il est répété jusqu'à l'obtention du nombre désiré de classes. Les seuils correspondent alors aux vallées de l'histogramme filtré (figure 4.14b). Dans notre cas, nous avons fixé le nombre désiré de classes à 4 pour obtenir les trois seuils. La figure 4.14 (a) présente un exemple où le fond est dans les mêmes gammes d'intensité et de teinte que le visage, Cette première impression peut être vérifiée sur l'image de la teinte (voir figure 4.9b) où l'on note une faible variation entre le fond et le visage. Dans ce cas, l'information de teinte n'est pas privilégiée et la stratégie de segmentation est principalement basée sur la luminance.

La figure 4.14 (c) montre l'histogramme de la Saturation de l'image 4.14 (b), qui est utilisé pour déterminer les paramètres de segmentation : S_0 S_T et S_L . Nous avons appliqué la méthode de Chang et nous avons obtenu le seuil maximal d'utilisation de la luminance S_L égal à 62, le seuil minimal d'utilisation de la teinte S_T égal à 240 et $S_0=176$. Enfin, pour obtenir une pente assez forte nous avons choisi $\beta = 0.01$ pour $\alpha(S_0) = 0.5$, et avons fixé le seuil sur la profondeur à 25.

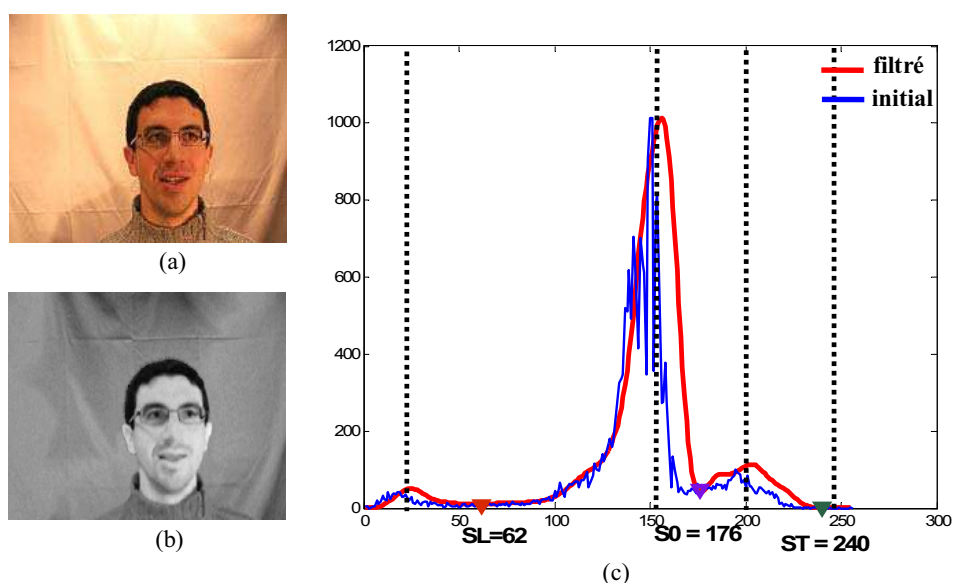


Figure 4. 14. (a) image originale (b) image de saturation (c) histogrammes de la saturation et seuils obtenus avec la méthode de Cheng.

La figure 4.15 (a) présente un exemple où le fond est uni de couleur bleu sombre. Cette première impression peut être vérifiée sur l'image de la teinte (voir figure 4.9b) où l'on note une haute variation entre le fond et le visage. Dans ce cas, l'information de teinte est privilégiée et la stratégie de segmentation est principalement basée sur teinte et la luminance.

Nous avons obtenue le seuil maximal d'utilisation de la luminance S_L égale à 58, le seuil minimal d'utilisation de la teinte S_T égale à 166 et $S_0=133$ (voir figure 4.14 (c)). Enfin, pour obtenir une pente assez forte nous avons choisi $\beta = 0.45$ pour $\alpha(S_0) = 0.5$, et avons fixé le seuil sur la profondeur à 15.

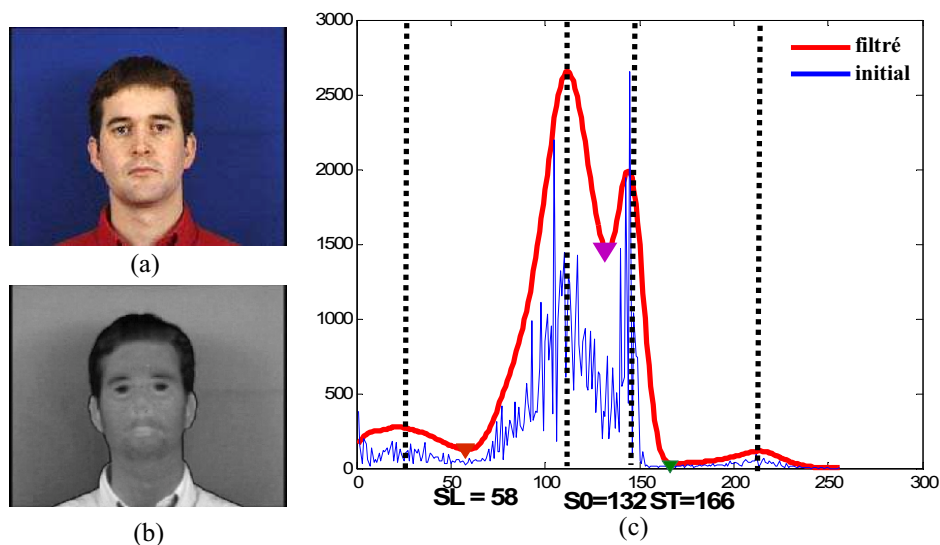


Figure 4. 15. (a) image originale (b) image de saturation (c) histogrammes de la saturation et seuils obtenus avec la méthode de Cheng.

4.4.5 Extraction de visage

L'algorithme de LPE modifié permet d'obtenir de grandes régions homogènes ce qui facilite l'interprétation de l'image. De plus le résultat de la segmentation est une image dans laquelle chaque pixel est affecté de l'étiquette (ou label) correspondant au numéro de la région à laquelle appartient le pixel dans l'image initiale. À partir de cette image "d'étiquettes" et de l'image originale il est possible de déterminer les divers attributs photométriques : niveau moyen, variance, indices texturaux, et géométriques : coordonnées du centre de gravité, axe d'inertie, de chaque région. Afin d'extraire le visage à partir de l'image segmentée, nous avons combiné deux informations : la taille et la teinte des régions obtenues. Une analyse des histogrammes de la teinte de 100 images de visages (ethniques, variation de l'illumination), nous a permis de définir l'intervalle d'appartenance D_T pour la teinte qui correspond à la couleur de la peau. Nous avons trouvé $D_T = [1 - 40]$. Pour chaque région segmentée, nous avons calculé la moyenne de la teinte. Parmi les régions de couleur appartenant au domaine « peau » (dont la teinte moyenne appartient à D_T) nous avons retenu celle qui a la taille maximale. Nous décidons que cette dernière correspond au visage selon la règle suivante :

$$\text{Région visage} \Leftrightarrow \text{Si } ((\text{Taille Région} = \max) \text{ ET } (\text{Teinte} \in D_T)) \quad (4.21)$$

Nous avons appliqué notre algorithme sur des images réelles qui présentent différentes conditions d'acquisition (ombre, éclairage variable, poses différentes). Les résultats obtenus (figures 4.16-b, 4.21-b, 4.22-b) illustrent le bon comportement de notre algorithme. En effet, celui-ci arrive à extraire la région du visage dans tous les cas de figure, ce qui démontre sa robustesse aux variations d'éclairage et de pose.

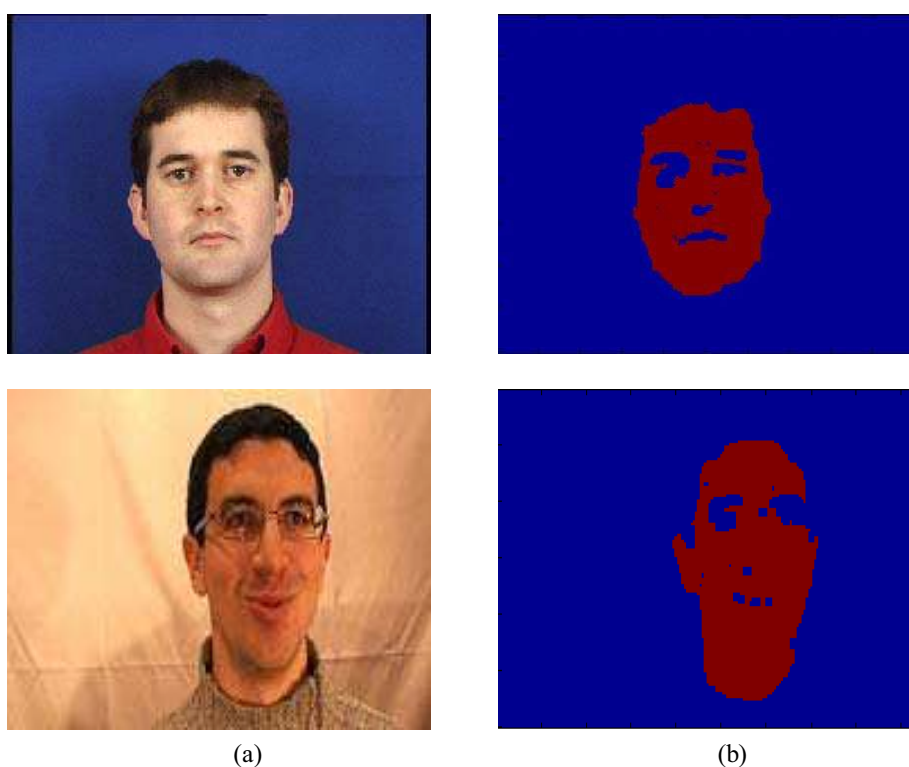


Figure 4. 16. (a) image originale, (b) résultats d'extraction du visage.

4.4.6 Extraction des régions caractéristiques du visage

Nous avons souligné dans la section 1.6.2 que l'extraction des régions caractéristiques du visage tel que les yeux, la bouche ou le nez, est une étape importante dans le processus de reconnaissance faciale. Nous avons donc développé une méthode pour l'extraction des zones caractéristiques du visage qui exploite les résultats obtenus par la segmentation. Elle utilise la géométrie du visage afin de classifier les régions candidates en trois régions caractéristiques : œil droit, œil gauche et bouche.

4.4.6.1 Modèle du visage

Plusieurs techniques efficaces de localisation des régions caractéristiques faciales qui utilisent la géométrie du visage ont été développées ces dernières années. Jeng et al. [Jen98] ont proposé une approche basée sur la configuration géométrique des régions faciales. Leur méthode permet de détecter des visages aussi bien en vue frontale qu'inclinée. Cependant elle échoue si la taille des visages est inférieure à 80 pixels. Lin et al [Lin99] ont développé un algorithme de détection de visages multiples dans des images à fond complexe. Ils supposent que, dans le cas d'une vue frontale du visage, les centres des yeux et le centre de la bouche forment un triangle isocèle, alors que dans le cas d'une vue de profil c'est le centre d'un oeil, l'orifice de l'oreille, et le centre de la région bouche qui forment un triangle. L'algorithme échoue quand les images sont trop foncées ou bien lorsque les yeux sont cachés par des cheveux. Nikolaidis et Pitas [Nik98] ont présenté une approche combinée qui utilise la transformée de Hough, le « template matching », les contours actifs et les propriétés projectives de la géométrie du visage. La transformée de Hough permet de détecter la région complète du visage, le template matching est utilisé pour identifier les régions caractéristiques du visage et les contours actifs servent à extraire les contours internes des régions faciales. Pour finir, les propriétés projectives de la géométrie de visage permettent de déterminer, avec précision, la pose du visage. Dans [Cri04], les auteurs utilisent les Modèles Actifs d'Apparence ou AAM (pour Active Appearance Model) pour prédire la localisation des caractéristiques faciales qui permettent de faire correspondre au mieux région faciale et modèle de visage préalablement appris. Duffner et Garcia [Duf05] ont proposé une méthode globale, basée sur les réseaux de neurones convolutionnels. Cette technique permet de détecter rapidement et précisément les régions caractéristiques du visage de manière robuste jusqu'à ± 60 dans le plan image et ± 30 profondeur, sur un fond complexe.

Les modèles géométriques de visage sont divers et variés. Chaque auteur construit son modèle en fonction des données dont il dispose. Toutefois, tous les modèles exploitent, d'une manière ou d'une autre, les distances discriminantes entre les régions caractéristiques d'un visage 2D (en vue frontale ou inclinée). Partant de ce constat, nous avons construit notre modèle géométrique de visage (frontal) en utilisant la distance entre les yeux Δ_{yeux} , qui représente la distance de référence à partir de laquelle les autres

paramètres du modèle sont déduits. Le modèle géométrique de visage que nous utilisons est présenté sur la figure 4.17. On détermine successivement :

- d) la droite L_1 passant par les yeux, puis la position des yeux sur cette droite,
- e) la droite L_2 passant par la bouche, puis la position de la bouche sur cette droite.

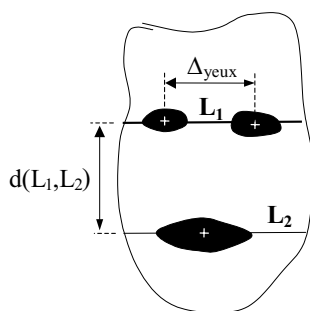


Figure 4. 17. Modèle géométrique du visage.

Les différentes étapes de l'algorithme qui permettent d'extraire ces trois régions du visage (les deux yeux et la bouche) sont détaillées ci-dessous.

4.4.6.2 Extraction des yeux

Tout d'abord nous commençons par déterminer la droite caractéristique L_1 de la région des yeux, pour cela nous parcourons les $1/3$ de la région du visage **binarisée**, notée I_M , de gauche à droite. Nous calculons pour chaque parcours le nombre de pixels égaux à zéro, la droite L_1 correspond au parcours ayant le maximum de pixels nuls. Ensuite, les régions les plus proches de la droite L_1 sont déterminées. Pour se faire, nous calculons la distance entre le centre de gravité de chaque région et la droite L_1 , les régions retenues sont celles dont la distance est inférieure à un seuil donné (régions R_i tels que : $d(R_i, L_1) < Th_1$, où $Th_1 = 8$ pixels). L'ensemble des régions retenues est classifié en deux classes, par la méthode des k-means. Chaque classe correspond à un œil (œil droit et œil gauche). Les régions de chaque classe sont ensuite fusionnées et le centre de gravité de chaque œil est calculé. Enfin le rectangle englobant chaque œil est dessiné pour illustrer le résultat de l'extraction.

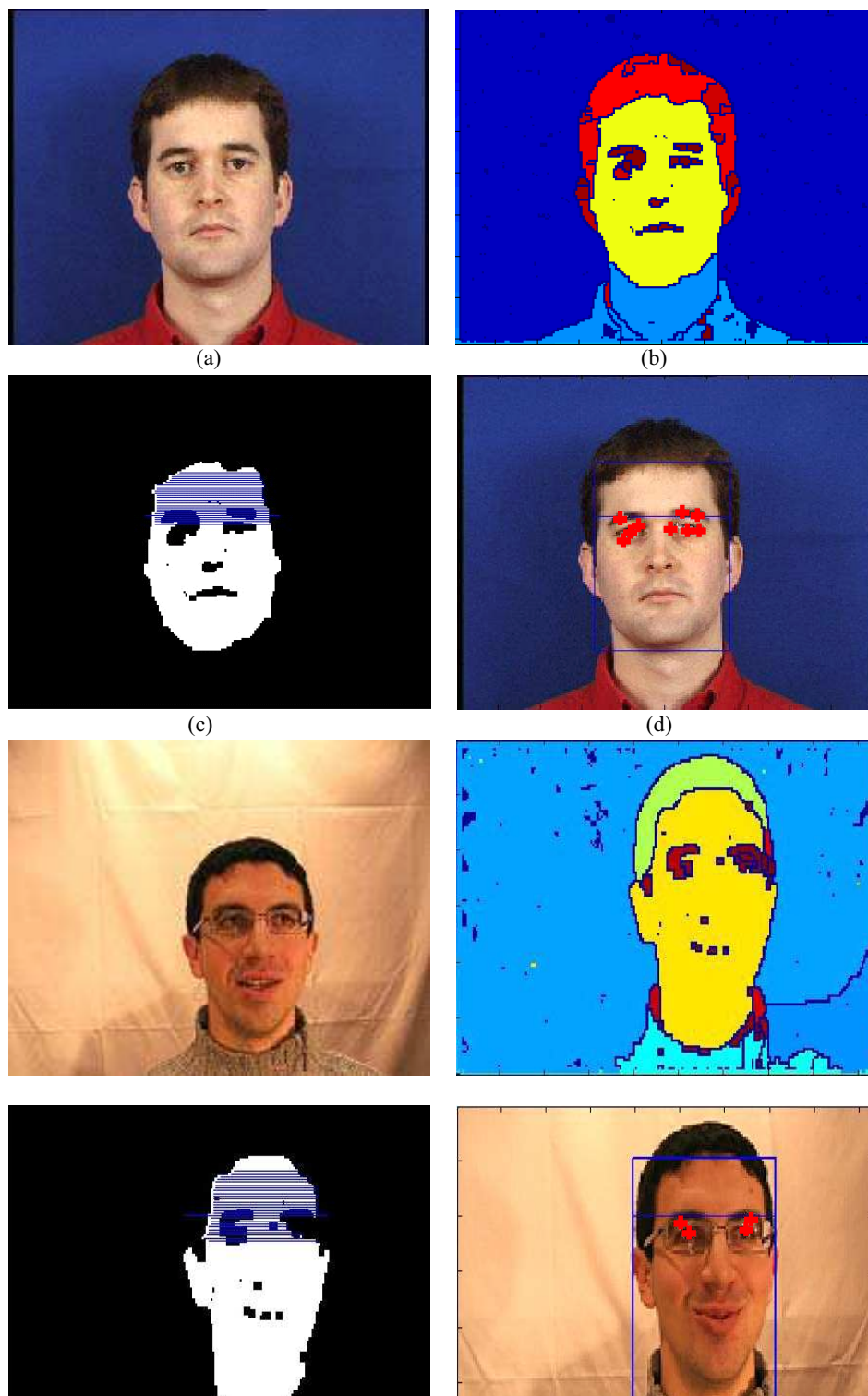


Figure 4. 18. (a) image originale, (b) visage segmenté I_s , (c) visage binarisé I_M , (d) les centres de régions yeux classifiées.

La figure 4.19, montre un exemple du processus de fusion des régions appartenant à la classe oeil droit.

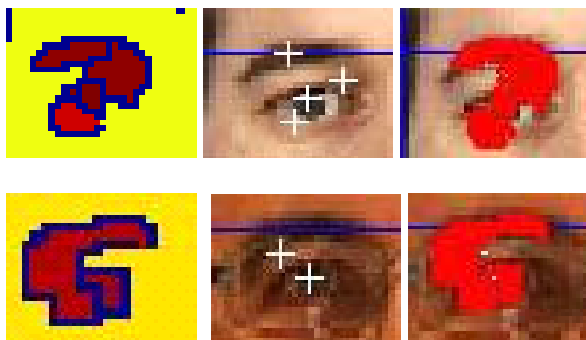


Figure 4. 19. Processus de fusion des régions appartenant à l'œil droit.

4.4.6.3 Extraction de la région de la bouche

De même, pour déterminer la région de la bouche nous commençons par déterminer la droite caractéristique L_2 avec équation suivante, (voir figure 4.20 (a)) :

$$D(L_1, L_2) = 5/6 * \Delta_{yeux} \quad (4.22)$$

Enfin, nous avons appliqué les mêmes traitements que précédemment pour déterminer la région de la bouche, à savoir :

- détermination des régions proches de la droite L_2 à un seuil donné (nous avons pris marge de 8 pixels)
- fusion des régions retenues (voir figure 4.20 (a))
- dessin du rectangle englobant la zone de la bouche (figure 3.20 (b)).

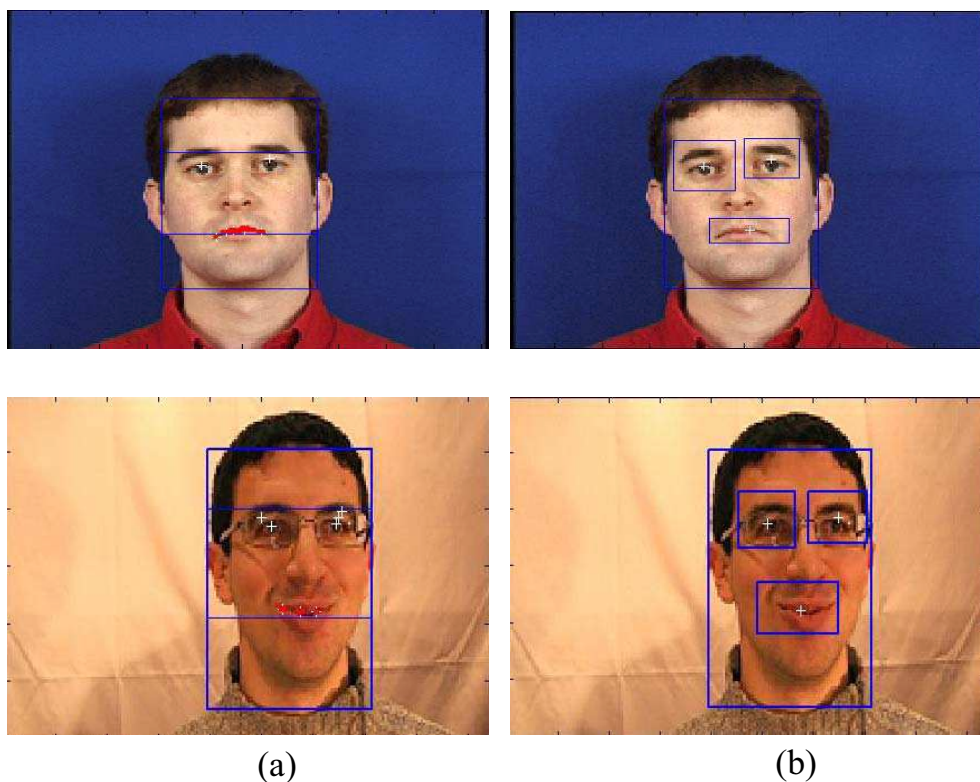


Figure 4. 20. Résultats d'extraction des régions caractéristiques.

4.4.6.4 Résultats visuels

Les figures 4.21 et 4.22 montrent quelques résultats obtenus avec notre algorithme d'extraction de régions caractéristiques sur des images de visages sous différentes conditions : changement d'illuminations et de pose, présence d'artefacts visuels (barbe et lunettes), et variation dans l'expression faciale. Les images originales proviennent des bases XM2VTS et IV².

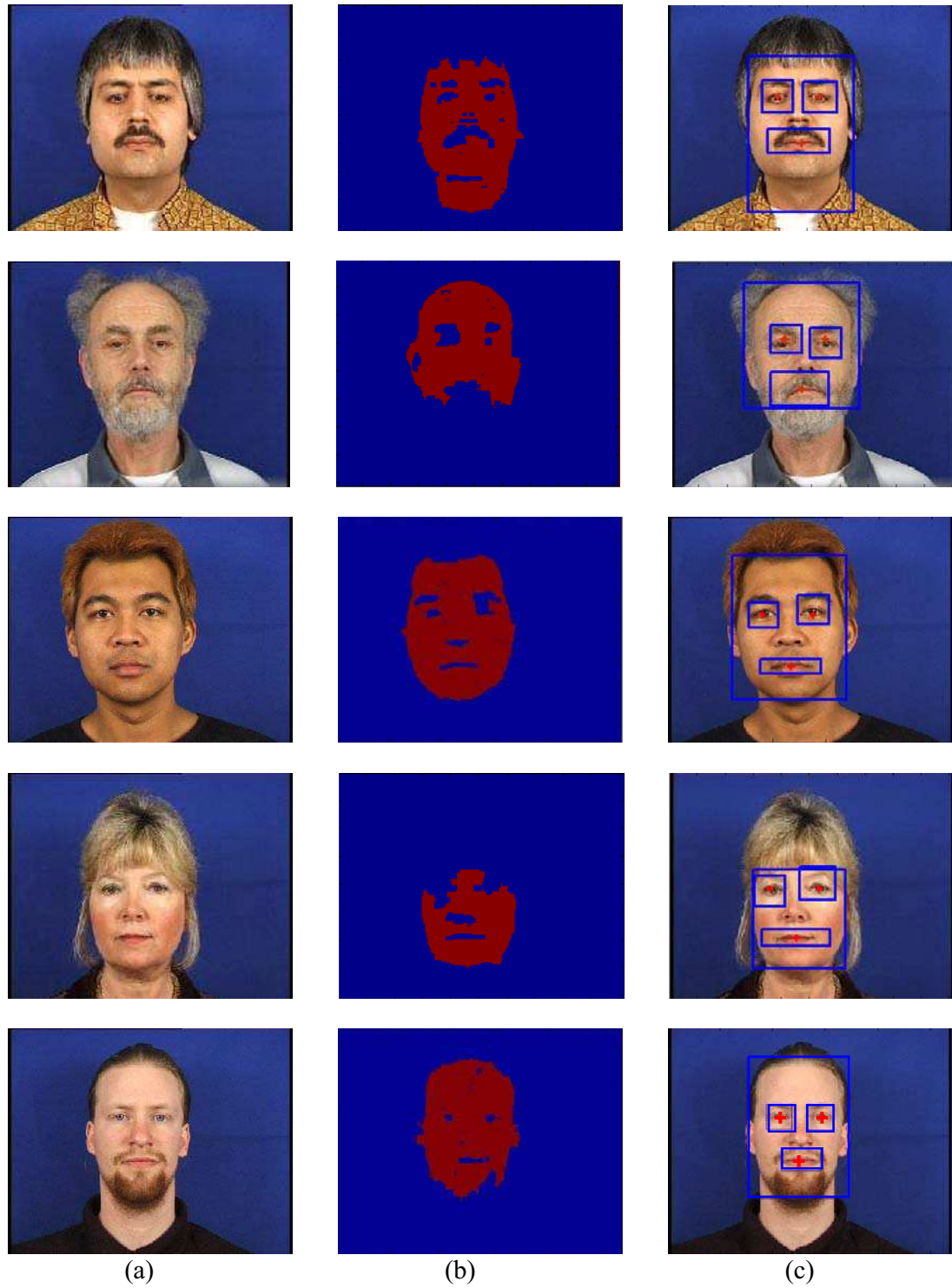
Base de données XM2VTS

Figure 4. 21. (a) image originale, (b) mask Visage (c) régions extraites

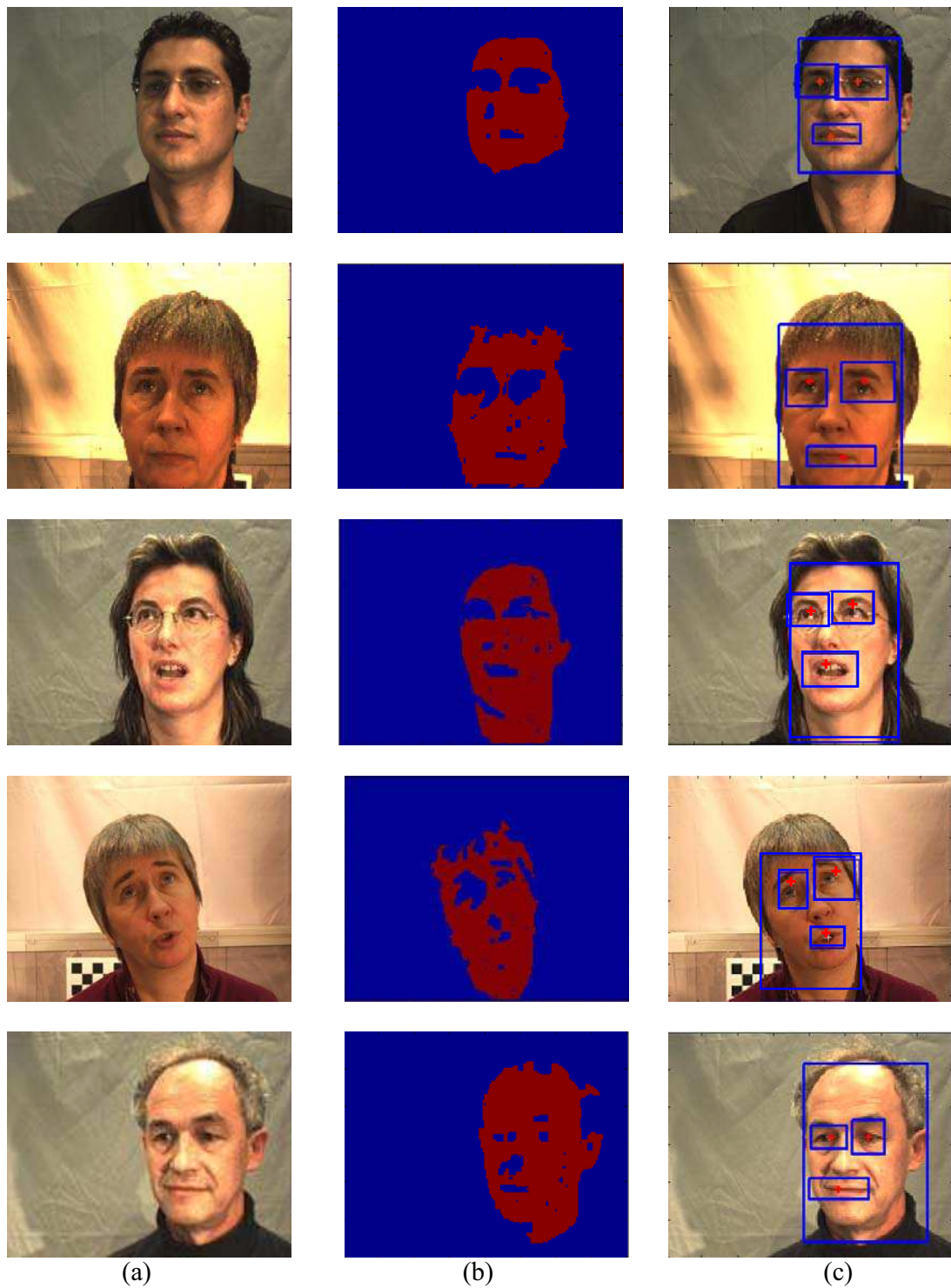
Base de données IV²

Figure 4. 22. (a) image originale (b) mask visage (c) régions extraites

4.4.6.5 Critères d'évaluations

L'extraction de visages est un sous domaine de l'analyse et de la compréhension d'images. Nous avons donc utilisé deux critères ayant trait à ce domaine que nous avons adaptés au cas de la localisation de visages, il s'agit de :

Détection positive

On appelle « détection positive » une fenêtre dans l'image qui, selon le détecteur, contient un objet caractéristique.

Détection négative

Réciproquement, on appelle « détection négative » une fenêtre dans l'image qui, selon le détecteur, ne contient pas d'objet caractéristique.

Le taux des bonnes détections positives (TBDP)

Le taux des détections d_p est le pourcentage des objets caractéristiques pour lesquels on a une détection positive dans une série d'images.

Le taux des mauvaises détections négatives (TMDN)

Le taux des mauvaises détections f_n est le pourcentage des objets caractéristiques pour lesquels on a une détection négative, dans une série d'images. La relation entre le taux des détections d_p et le taux des mauvaises détections négatives f_n est :

$$f_n = 1 - d_p \quad (4.23)$$

4.4.6.6 Expérimentation et Résultats

Dans notre cas, une détection positive de visage sera estimée de manière qualitative par la présence des yeux et de la bouche dans la fenêtre obtenue (voir figure 4.23-a). Pour les yeux on considéra qu'un œil est convenablement détecté si la pupille est présente dans la fenêtre. Pour la bouche, le critère est exigeant car on veut disposer de l'image de la bouche entière dans la fenêtre.(voir figure 4.23).

Les images suivantes montrent des exemples de mauvaises et bonnes détections pour le visage, l'œil droit, l'œil gauche et la bouche.

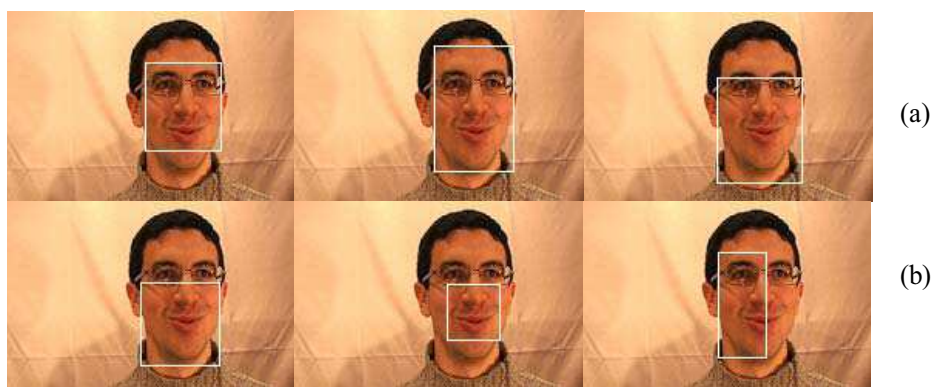


Figure 4. 23. (a) les bonnes détections (b) les mauvaises détections.

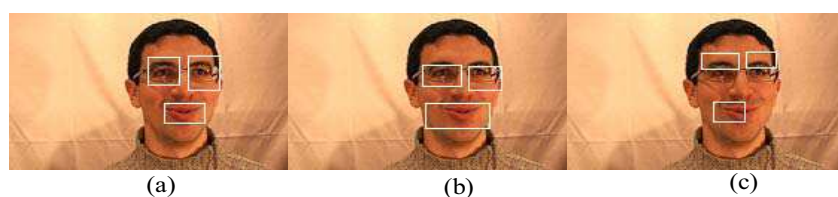


Figure 4. 24. (a,b) les bonnes détections (c) les mauvaises détections.

Nous avons appliqué notre approche sur 200 images de taille (256*256) de la base XM2VTS présentant différentes variations d'apparence : barbe, lunette, variation d'illumination.

XM2VTS /200	TBDP
visage	99%
bouche	97%
yeux	99%

Tableau 4. 2. Taux des bonnes détections positives sur la base XM2VTS.

Par ailleurs, nous avons aussi appliqué notre approche sur une cinquantaine d'images de la base IV², acquises par une paire stéréoscopique de résolution moyenne. Ces images présentent des variations significatives d'éclairage (ombrage, illumination) de pose et dans les expressions faciales.

IV ² /500	TBDP
visage	99%
bouche	98%
yeux	99%

Tableau 4. 3. Taux des bonnes détections positives sur la base IV².

4.4.7 Extraction et Normalisation des régions pour la vérification de visage

Nous avons appliqué notre méthode d'extraction de caractéristiques sur les 500 images de la base IV^2 de la première phase d'évaluation. Pour chaque image, nous avons extrait le visage ainsi que les positions des yeux et de la bouche. Ces données brutes ne sont pas directement exploitables par le module de reconnaissance du visage, un premier traitement est nécessaire (voir figure 4.25-a), il consiste à normaliser toutes les images de visage, en utilisant les coordonnées des centres des yeux (voir figure 4.25-b). Enfin, pour extraire les différentes régions d'intérêt nous avons utilisé le centre des yeux ainsi que la position de la bouche (figure 4.25-c). Les paramètres $y_0 = 15$ pixels et $b_0 = 20$ pixels (voir figure 4.25-b) définissent les hauteurs des régions yeux et bouche respectivement, ils sont fixés empiriquement. La région du nez correspond à la différence entre la région visage normalisé et les régions bouche et yeux.

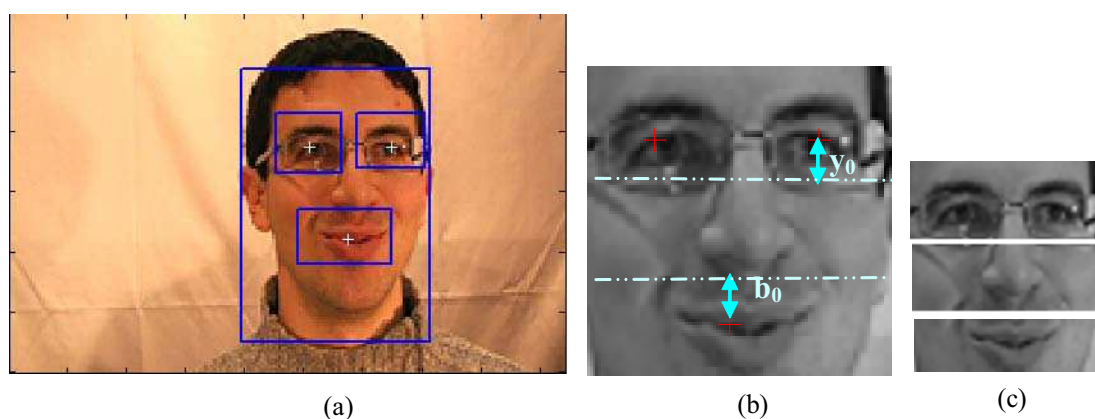


Figure 4. 25. (a) Les caractéristiques extraites avec notre méthode (b) visages normalisés (c) les régions extraites.

4.4.8 Conclusion

Nous avons proposé dans ce chapitre une méthode de segmentation basée sur la ligne de partage des eaux qui s'applique aux images couleurs afin d'extraire la région du visage. Elle est basée sur le calcul du gradient TLS et prend en compte la pertinence de la teinte lors de la fusion des bassins versants. Les résultats obtenus démontrent la robustesse au bruit et aux variations de l'illumination. L'avantage de cette méthode est qu'elle utilise différentes stratégies de segmentation en fonction de la valeur de la saturation (luminance, luminance-teinte, teinte). Par ailleurs, nous avons proposé un modèle géométrique simple et efficace du visage qui nous a permis d'extraire les trois régions caractéristiques du visage, à savoir : les deux yeux et la bouche. Nous avons évalué les performances de notre approche d'extraction sur les deux bases de données XM2VTS et IV². Bien que notre critère d'évaluation soit de nature qualitative et fondée uniquement sur la présence des composants essentiels du visage (yeux, nez et bouche), il nous a permis de valider l'efficacité de notre solution et sa relative robustesse aux différentes conditions d'acquisition. Cette approche sera utilisée pour l'extraction automatique des régions du visage dans les méthodes de reconnaissance 2D que nous allons détailler dans le chapitre suivant.

Chapitre 5

Reconnaissance de visage 2D/3D

5.1 Introduction

Nous avons décrit dans le chapitre 2 (section 2.3) plusieurs méthodes de reconnaissance 2D/3D de visage. La reconnaissance du visage 2D a atteint un haut niveau de performances, même avec l'utilisation d'un seul exemple d'apprentissage. Bien que le dernier rapport FRGC indique que de meilleures performances de reconnaissance automatique des visages ont été obtenues par des systèmes de vision plutôt que par des humains, ces résultats font référence à des conditions spécifiques (variations contrôlées). Dans les conditions actuelles des travaux sur la reconnaissance faciale, le système visuel humain reste encore le plus robuste face aux diverses variations pouvant altérer le processus d'identification : changement des conditions d'éclairage, variations de l'expression faciale, et/ou de l'orientation du visage, modifications de l'apparence du visage à travers la présence ou l'absence de lunettes, barbe, maquillage. La résolution de ces verrous rendrait les techniques de reconnaissance 2D du visage beaucoup plus attrayantes pour les besoins d'authentification et/ou de vérification d'identité. Par ailleurs, nous avons souligné que la reconnaissance faciale 3D a été identifiée comme étant une solution intéressante aux problèmes cités ci-dessus. Cependant, ces techniques n'ont pas

encore atteint leur maturité. En effet, la plupart des bases de données de visages 3D, utilisées pour le « Benchmarking », ne sont pas encore standardisées et leurs dimensions varient considérablement. De plus, le protocole expérimental et la métrique utilisés, pour évaluer les performances de ces techniques, sont très différents d'un travail de recherche à un autre.

Bien que la reconnaissance 3D du visage soit une voie émergente, il est nécessaire de pouvoir comparer ses performances avec celles des techniques classiques dans un cadre contrôlé où elle serait soumise au même protocole d'évaluation sur une grande base de données. Ce besoin a incité les principaux fournisseurs de données biométriques à réaliser des tests sur des données 3D de visages. Ainsi, la dernière évaluation du FRVT (FRVT 2006 (<http://www.frvt.org/>) ne s'est pas limitée aux techniques de reconnaissance 2D du visage mais inclus aussi les techniques de reconnaissance 3D de visage (sur la base FRGC). Afin de répondre à ces nouveaux besoins en matière d'évaluation des techniques de reconnaissance faciale 2D et 3D, il est nécessaire d'avoir une grande et riche base de données contenant des visages acquis sous différentes conditions : variations d'expressions, de poses et d'éclairage, etc. Dans notre cas, nous avons utilisé la base de données IV² afin d'évaluer les différentes approches élaborées pour la reconnaissance 2D/3D de visage.

Ce chapitre est divisé en trois parties : dans la première partie, nous décrivons en détail notre technique d'authentification 2D ; nous donnons aussi les résultats d'évaluations obtenus sur les données de la base IV², après détection de visage et extraction des caractéristiques par la méthode décrite au chapitre 4. Dans la deuxième partie nous présentons notre approche 3D de reconnaissance basée sur les mesures anthropométriques du visage. Enfin, dans la dernière partie de ce chapitre nous exposons une approche hybride de reconnaissance du visage qui combine les approches 2D/3D.

5.2 Authentification de visage 2D

De nombreuses techniques 2D ont été développées ces dernières années, nous les avons passées en revue dans le chapitre 2. Parmi elles, l'Analyse en Composantes Principales (ACP) est une technique particulièrement prisée par les chercheurs de la communauté de la

biométrie. Elle est utilisée soit de façon globale sur toute l'image du visage, soit de façon modulaire sur les différentes régions faciales. De plus, plusieurs extensions de l'ACP ont aussi été proposées et utilisées pour la reconnaissance faciale. Par ailleurs, nous avons souligné l'influence du nombre d'exemples d'apprentissage par personne sur les performances de la reconnaissance, et avons indiqué que l'ACP reste une méthode efficace et simple pour gérer ce type de problème. C'est pour toutes ces raisons que nous avons opté pour l'ACP dans notre approche 2D de reconnaissance du visage.

Nous commençons d'abord par détailler le principe des méthodes Eigenface et Eigenface modulaire.

5.2.1 Eigenface globale

La technique Eigenface globale est une méthode très utilisée pour la reconnaissance de visage. Sa popularité est due à sa nature relativement simple, son fondement mathématique fort et ses résultats bons.

Soit $I(x,y)$ une image de dimension $N \times N$, et x un vecteur de dimension $L = N^2 \times 1$. Soit S l'ensemble de vecteurs représentant les images de visage dans la base de données de dimension M . S est défini par :

$$S = \{x_1, x_2, \dots, x_M\} \quad (5.1)$$

La méthode Eigenface est constituée de deux processus ou phases : un processus d'apprentissage et un processus de reconnaissance. Ces derniers peuvent être résumés comme suit :

Processus d'apprentissage

- Déterminer la moyenne des vecteurs images, tel quel :

$$\Psi = \frac{1}{M} \sum_{n=1}^M x_n \quad (5.2)$$

- Calculer la différence entre les vecteurs image d'entrée x_i et l'image moyenne :

$$\Phi_i = x_i - \Psi \quad 1 \leq i \leq M \quad (5.3)$$

- Déterminer un ensemble de M vecteurs orthogonaux u_n qui représentent le mieux la distribution des données. Le $k^{\text{ème}}$ vecteur u_k est choisi telle que :

$$\lambda_k = \frac{1}{M} \sum_{n=1}^M (u_k^T \Phi_n)^2 \quad (5.4)$$

soit maximale. Où u_k, λ_k sont respectivement les vecteurs propres et les valeurs propres de la matrice de covariance C définie par :

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T \quad (5.5)$$

La figure (5.1) montre les cinq premières Eigenfaces associés aux plus grandes valeurs propres, et calculées sur la base d'apprentissage IV^2 (pour 25 personnes).



Figure 5. 1. Exemple de décomposition en Eigenfaces.

Processus de reconnaissance

Plusieurs travaux [Tur91] [Mar01] ont démontré qu'en pratique, un nombre $M' < M$ de Eigenfaces est généralement suffisant pour identifier efficacement les visages. Les M' visages propres considérés correspondent aux plus grandes valeur propres λ_k car elles encodent les directions dans lesquelles les variations sont les plus marquées. Les premiers visages propres représentent habituellement les différences d'éclairage ainsi que les personnes portant des lunettes ou une barbe.

Le processus de reconnaissance se subdivise en trois étapes : l'image d'entrée $I(N \times N)$ contenant le visage à identifier est transformée en un vecteur X de dimension $(N^2 \times 1)$ puis comparé avec le vecteur moyen Ψ défini par l'éq. (5.2). La différence entre le vecteur X et

le vecteur moyen Ψ est ensuite multipliée par l'ensemble des vecteurs propres u_k pour générer le vecteur de poids Ω_X^T , tel que :

$$\varpi_k = u_k^T (X - \Psi) \quad k \in \{1, \dots, M'\} \quad (5.6)$$

$$\Omega_X^T = [\varpi_1, \varpi_2, \dots, \varpi_{M'}] \quad (5.7)$$

La classe de visages qui fournit la meilleure description de l'image d'entrée est déterminée en minimisant la mesure de similarité entre le poids Ω_X^T et les poids de la base d'apprentissage $\Omega_{x_k}^T \quad k = 1, \dots, M'$.

5.2.2 Eigenface modulaire

Afin de surmonter les problèmes des variations d'expressions et d'éclairage, Pentland et al [Pen94] ont proposé d'appliquer l'ACP de manière modulaire sur les différentes régions faciales (voir section 2.3). La motivation principale de cette approche est que les différentes régions faciales ne sont pas affectées de la même manière par les différentes sources de variabilité. Par exemple, les expressions faciales affectent plus la bouche que les autres régions faciales. Ainsi, par rapport à la méthode Eigenface globale, la méthode Eigenface modulaire nécessite une étape de prétraitement supplémentaire qui consiste à localiser les régions d'intérêt à l'intérieur du visage lors des phases d'apprentissage et de reconnaissance. La précision du module de détection des caractéristiques de visage est donc cruciale pour les performances de cette méthode.

Phase d'apprentissage : durant cette phase, les régions du visage sont extraites puis classées en un ensemble de régions. L'ACP correspondante à chaque région est ensuite calculée en utilisant la méthode d'apprentissage présentée dans la sous-section 5.2.1.

Phase de reconnaissance : pour chaque image d'entrée, les caractéristiques de visage sont tout d'abord extraites puis utilisées indépendamment par la procédure de reconnaissance basée Eigenfaces. Les distances ainsi obtenues de l'ensemble des sous-images sont ensuite fusionnées pour générer un résultat global de classification.

5.2.3 Tests et Evaluations

Afin d'étudier l'influence des différentes régions du visage (yeux, nez, bouche,...) sur la reconnaissance d'un individu, nous proposons une approche modulaire qui évalue l'ACP sur un ou plusieurs classifieurs. Chaque classifieur correspond à une région particulière du visage. Le but est de démontrer que cette combinaison ou fusion de classifieurs peut nettement améliorer le taux de reconnaissance.

Nous avons ainsi mené deux évaluations pour analyser la précision du module de détection du visage :

- La première consiste à évaluer les méthodes Eigenface globale et modulaire sur les régions du visage extraites manuellement.
- La deuxième permet d'évaluer ces mêmes méthodes sur les régions extraites, cette fois-ci, avec notre méthode de détection automatique de visage et des régions caractéristiques.

Afin d'obtenir des éléments de comparaison, nous avons utilisé le protocole d'évaluation de la phase préliminaire de IV² (voir chapitre 3), à savoir : 50 personnes, 10 images par personnes, ce qui fait une base de 500 images. Nous avons séparé les individus en deux groupes de 25 : le premier groupe est utilisé pour le test et le deuxième groupe pour construire les classifieurs ACP (avec un seul exemple d'apprentissage par personne). A partir des 250 images (25×10) de la base de test, nous avons effectué 500 comparaisons deux à deux dont 250 comparaisons intra classes et 250 comparaisons inter-classes (voir chapitre 3). Le taux de vérification est mesuré avec le taux EER ainsi que l'intervalle de confiance (IC) associé à 90%.

On obtient l'intervalle de confiance [Ben03] CI_{FAR} sur le taux de fausse acceptation via l'équation (1.3), :

$$CI_{FAR} = FAR \pm \alpha \cdot \sqrt{\frac{1}{NI} \cdot FAR(1 - FAR)} \quad (5.8)$$

où $\alpha = 1,64$ décrit un intervalle de confiance à 90%, $\alpha = 1.960$ un intervalle de confiance à 95% et $\alpha = 2.576$ un intervalle de confiance à 99%. On obtient pour les faux rejets :

$$CI_{FRR} = FRR \pm \alpha \cdot \sqrt{\frac{1}{NC} \cdot FRR(1-FRR)} \quad (5.9)$$

où NI et NC sont les nombres d'accès imposteur et client respectivement. On calcule l'intervalle de confiance du EER à partir de l'équation(5.10), tmpEER est indice EER où la valeur FAR = FRR

$$CIEER = EER - 100 * (CIFAR(tmpEER) + CIFRR(tmpEER)) / 2 \quad (5.10)$$

5.2.3.1 Evaluation 1 : Extraction manuelle

Pour ce test, Nous avons tout d'abord mené trois expériences différentes :

- La première consiste à évaluer l'influence de la résolution des images de visage sur les performances du système, elle permet de déterminer les résolutions optimales.
- La deuxième consiste à évaluer l'influence des mesures de similarité sur les performances du système.
- La troisième expérience permet d'évaluer les performances de la combinaison de plusieurs classifieurs (visage, yeux, nez, bouche).

Expérience 1 : choix de la résolution image

Nous avons redimensionné les images de visage normalisées de taille (220×200) en différentes résolutions {15×13, 75×65, 100×90, 150×130}. Pour chaque résolution, nous avons effectué une comparaison entre les signatures (projections) associées aux vecteurs image $X = [X_1, \dots, X_n]^T$ et $Y = [Y_1, \dots, Y_n]^T$ selon la distance L_1 , appelée distance de Manhattan :

$$d_{L_1}(X, Y) = \sum_{i=1}^n |X_i - Y_i| \quad (5.11)$$

Le tableau 5.1 montre les résultats du taux EER obtenus pour les différentes valeurs de résolution ainsi que l'intervalle de confiance et les temps de calcul. Nous remarquons qu'une diminution de la résolution, dans certain cas, n'engendre pas une baisse

significative du EER. Nous constatons aussi que le EER le plus faible est égal à 0.077 ± 0.029 , il correspond à une résolution de 100×90 . Nous avons donc choisi de prendre une résolution de 75×65 qui donne un EER égal à 0.079 ± 0.03 , pour avoir un meilleur compromis entre temps de calcul et précision.

Visage 2D	15×13	75×65	100×90	150×130	220×200
EER	0.18	0.079	0.077	0.082	0.081
IC	±0.04	±0.029	±0.028	±0.029	±0.029
Temps	21s	32s	46s	76s	137s

Tableau 5. 1 Les résultats des EER, l'intervalle de confiance à 90% associé et le temps exécution pour l'expérience 1.

Expérience 2 : choix des mesures de similarité

La deuxième expérience est menée sur des images ayant une résolution de 75×65 . Nous avons comparé trois distances : la distance L_1 (équation 5.8), la distance L_2 et la distance Cosinus telles que :

- la distance L_2 , appelée distance euclidienne, est définie par :

$$d_{L_2}(X, Y) = \sum_{i=1}^n (X_i - Y_i)^2 \quad (5.12)$$

- la mesure de similarité d'angle négatif (ou distance cosinus), entre les vecteurs X_i et Y_i est donnée par :

$$d_{\cos}(X, Y) = \frac{\sum_{i=1}^n X_i Y_i}{\sqrt{\sum_{i=1}^n X_i^2 \sum_{i=1}^n Y_i^2}} \quad (5.13)$$

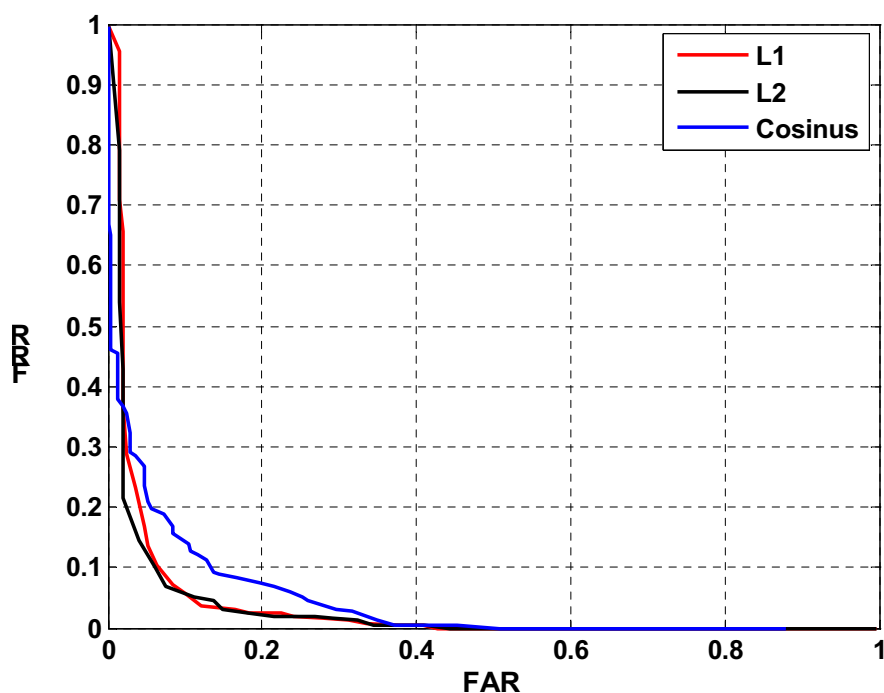


Figure 5. 2. Courbes DET : les résultats sur la deuxième expérience.

La figure 5.2 montre la courbe DET pour les trois distances. Nous remarquons que les distances L_1 et L_2 donnent de meilleurs résultats que la distance Cosinus. Par ailleurs, les résultats des distances L_1 et L_2 sont très similaires ($EER_{L_1} = 0.073 \pm 0.027$ et $EER_{L_2} = 0.079 \pm 0.028$). Nous avons donc choisi d'utiliser la distance L_1 pour le reste de l'évaluation des performances.

Expérience 3 : combinaison des classifieurs

Nous avons extrait manuellement les régions caractéristiques en découpant le visage normalisée en trois parties distinctes : yeux, nez, et bouche. Nous avons appliqué une ACP sur chacune des quatre régions du visage, à savoir: visage global, yeux, nez et bouche. Pour la méthode Eigenface modulaire, les scores de la classification obtenus sur chacune des régions (visage, yeux, bouche et nez) sont fusionnés. Pour étudier l'influence de chaque région sur les performances de la reconnaissance, nous avons affecté à chaque région un poids différent. Les régions faciales à partir desquelles sont construits les classifieurs sont illustrées dans la figure 5.3. Le classifieur 1 est construit à partir de l'image globale du visage de taille 75×65 . Le classifieur 2 est construit à partir d'une région de taille 50×65 contenant les yeux et les sourcils, c'est la région la plus stable du

visage. Le troisième classifieur contient la région du nez de taille 35×65 . Enfin, le quatrième classifieur contient la région de la bouche de taille 35×65 . Chaque classifieur est préalablement construit en appliquant une ACP sur la base d'apprentissage (25 personnes) propre à chaque région. Ceci permet d'obtenir une matrice par région avec un nombre de vecteurs propres associés.

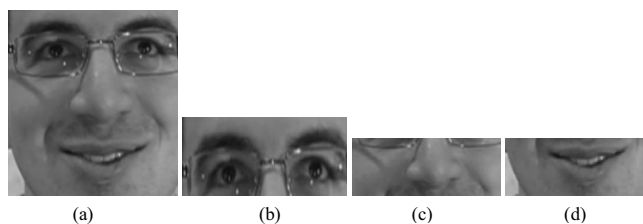


Figure 5. 3. Régions faciales utilisées pour construire les quatre classifieurs.

Pour étudier l'influence de chaque région sur la classification, nous avons calculé son taux EER. La figure (5.4) montre les résultats obtenus pour les différentes régions du visage.

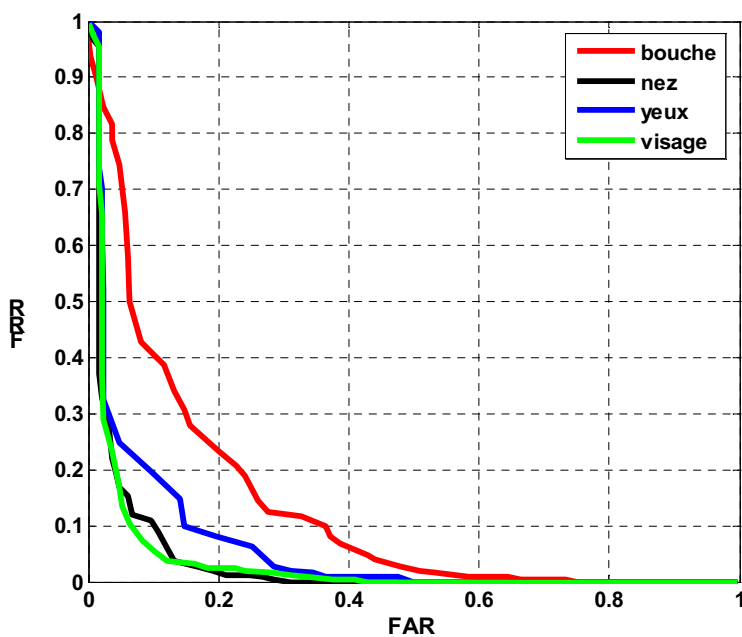


Figure 5. 4. Courbe DET : des différents classifieurs.

Nous remarquons que la région bouche donne les plus mauvais résultats ($ERR = 0.21 \pm 0.042$) en comparaison avec les régions yeux ($ERR = 0.14 \pm 0.035$), nez ($ERR = 0.10 \pm 0.031$) et visage ($ERR = 0.079 \pm 0.0281$).

La figure 5.5 montre la stratégie de la fusion des scores que nous avons adoptée. La classification de chaque image requête se fait de la manière suivante :

- Extraction des régions.
- Calculer les signatures de l'image requête.
- Comparer ces signatures avec les signatures du classifieur lié à la région utilisant la distance L_1 . Nous obtenons ainsi quatre scores s_1, s_2, s_3, s_4 (un score par région).
- Les quatre scores sont ensuite fusionnés pour calculer un score final de classification.

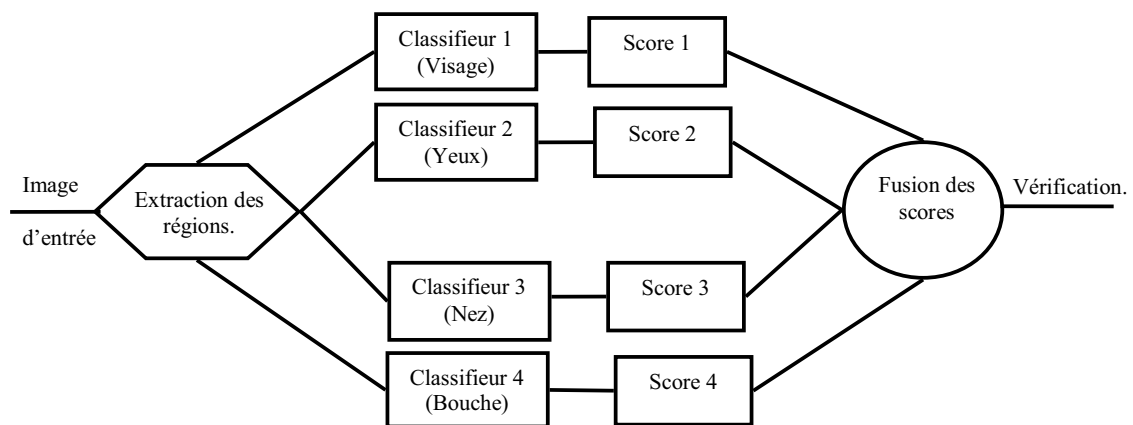


Figure 5. 5. Fusion des scores. La classification se fait grâce à des règles permettant de combiner les sorties des quatre classifieurs.

Pour trouver le meilleur compromis entre les régions nous avons testé plusieurs combinaisons de fusion. Pour cela, nous avons utilisée deux méthodes de fusion : la fusion par somme directe et la fusion par somme pondérée.

- La fusion des scores par une somme directe est définie par:

- $S_{ynv} = S_{yeux} + S_{nez} + S_{global}$
- $S_{yn} = S_{yeux} + S_{nez}$
- $S_{ynb} = S_{yeux} + S_{nez} + S_{bouche}$

La figure 5.6 montre les résultats obtenus pour ces différentes combinaisons. Elle présente les courbes DET des fusions de scores. Nous remarquons que la fusion S_{ynb} (yeux, nez et bouche) engendre le plus mauvais taux de $EER = 0.072 \pm 0.026$. Par contre, la fusion des scores des régions yeux et nez donne le taux le plus faible ($EER = 0.046 \pm 0.021$). Nous avons aussi appliqué la fusion S_{ynv} décrite dans [Pen94] qui combine les régions des yeux et du nez avec la région globale du visage, le taux ERR obtenu est égal à 0.050 ± 0.022 .

- la fusion par somme pondérée (Matcher Weighting [Sne05]) des différents scores: les pondérations sont assignées à chaque classifieur en se basant sur le taux EER qui lui est associé.

En notant le EER d'un classifieur m comme r_m , $m = 1, 2, \dots, M$. où M est le nombre total de classifieurs. La pondération α_m associée au classifieur m est donnée par :

$$\alpha_m = \frac{\left(1 / \sum_{m=1}^M \frac{1}{r_m}\right)}{r_m} \quad (5.14)$$

Notons que $\forall m, \sum_{m=1}^M \alpha_m = 1$ et les pondérations sont inversement proportionnelles aux erreurs correspondantes et sont par conséquent plus grandes pour les classifieurs les plus précis. La quantité s_m représente le score normalisé pour le classificateur m . La fusion des scores notée S_p est calculée comme suit :

$$S_p = \sum_{m=1}^M \alpha_m s_m \quad (5.15)$$

A partir des valeurs des EER de chaque classifieurs (yeux, nez, bouche) trouvées précédemment nous avons calculé les pondération α_m ($m=1,2,3$) en utilisant l'équation (5.14), nous avons obtenu: $\alpha_{yeux}=0.4461$, $\alpha_{nez} = 0.341$ et $\alpha_{bouche}= 0.212$. Le score final $S_p = 0.4461 * S_{yeux} + 0.341 * S_{nez} + 0.212 * S_{bouche}$, avec un taux $EER = 0.57 \pm 0.024$ pour :

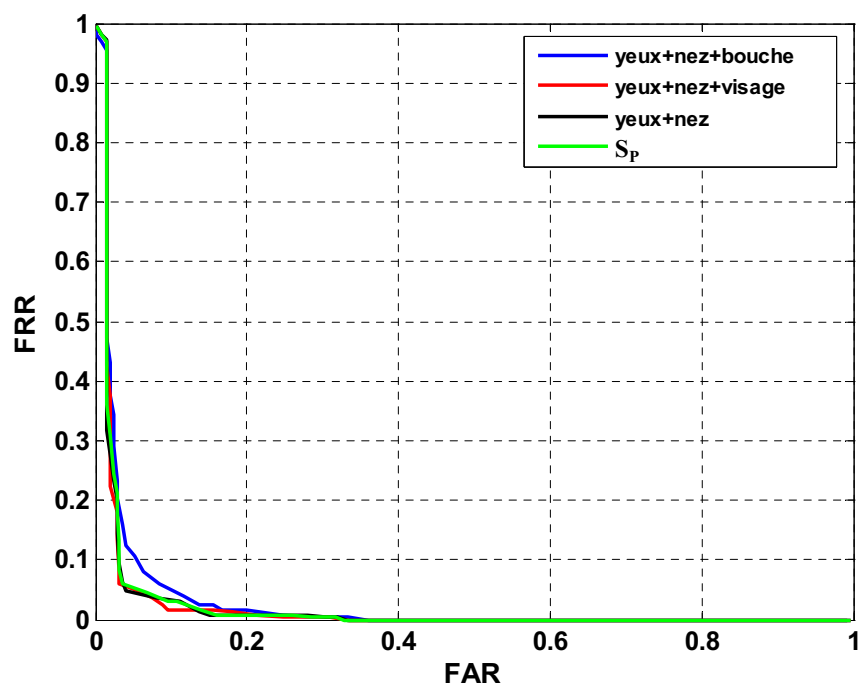


Figure 5. 6. Les courbes DET des différentes combinaisons de classifieurs.

En conclusion, cette expérience a démontré qu'avec l'extraction manuelle des régions de visage, la fusion nez et yeux donne le meilleur taux EER, soit 0.046 ± 0.021 .

5.2.3.2 Evaluation 2 : Extraction automatique

Dans cette deuxième évaluation, nous allons étudier l'influence du module de détection automatique de visage et de ses caractéristiques (algorithme présenté dans le chapitre 4) sur le système de la reconnaissance faciale.

Nous avons évalué les méthodes de reconnaissance 2D globale et modulaire en utilisant les régions images extraites automatiquement. Nous avons comparé les résultats obtenus avec ceux de l'évaluation 1 (extraction manuelle). La figure 5.7 montre les différentes courbes DET des régions du visage. Nous remarquons que dans tous les cas, le taux EER pour les régions extraites automatiquement est plus faible. Ainsi, par exemple, le taux EER pour la région visage extraite automatiquement est inférieur de 2% à celui de la région visage extraite manuellement. Pour la région bouche cette baisse est de 0.7%.

Par ailleurs, nous avons comparé les résultats de la fusion (voir figure 5.8) des classifieurs dans le cas de l'extraction automatique de régions, nous avons constaté que le taux a baissé d'environ 1%. Le taux le plus faible $EER = 0.044 \pm 0.021$ correspond dans ce cas à la fusion pondérée :

$$S_P = 0.4511 * S_{Nez} + 0.2911 * S_{Yeux} + 0.2578 * S_{Bouche} \quad (5.16)$$

Ce résultat, démontre que même si la région bouche possède un taux de vérification faible, surtout lorsque la base de données contient des variations d'expressions, il ne faut pas l'exclure de la stratégie de fusion, il faut juste lui affecter une pondération plus faible.

Il est vrai que cette première phase de test du projet IV² ne donne pas les résultats définitifs, mais elle nous a permis de mettre en avant les performances de la fusion des régions caractéristiques pour la reconnaissance faciale, et particulièrement la fusion pondérées des régions du nez, des yeux et de la bouche. Nous avons donc retenu cette approche de fusion pour la méthode de reconnaissance hybride 2D+3D qui sera décrite dans la troisième partie de ce chapitre.

Région	EER ± IC Evaluation 2 : Extraction automatique	EER ± IC Evaluation 1 : Extraction manuelle
Yeux	0.120±0.0338	0.14±0.035
Nez	0.079± 0.0281	0.10±0.031
Bouche	0.14±0.035	0.21±0.042
Visage	0.059± 0.0245	0.079± 0.0281

Tableau 5. 2. Comparaison entre les taux EER obtenus dans les cas de l'extraction automatique et manuelle des régions du visage.

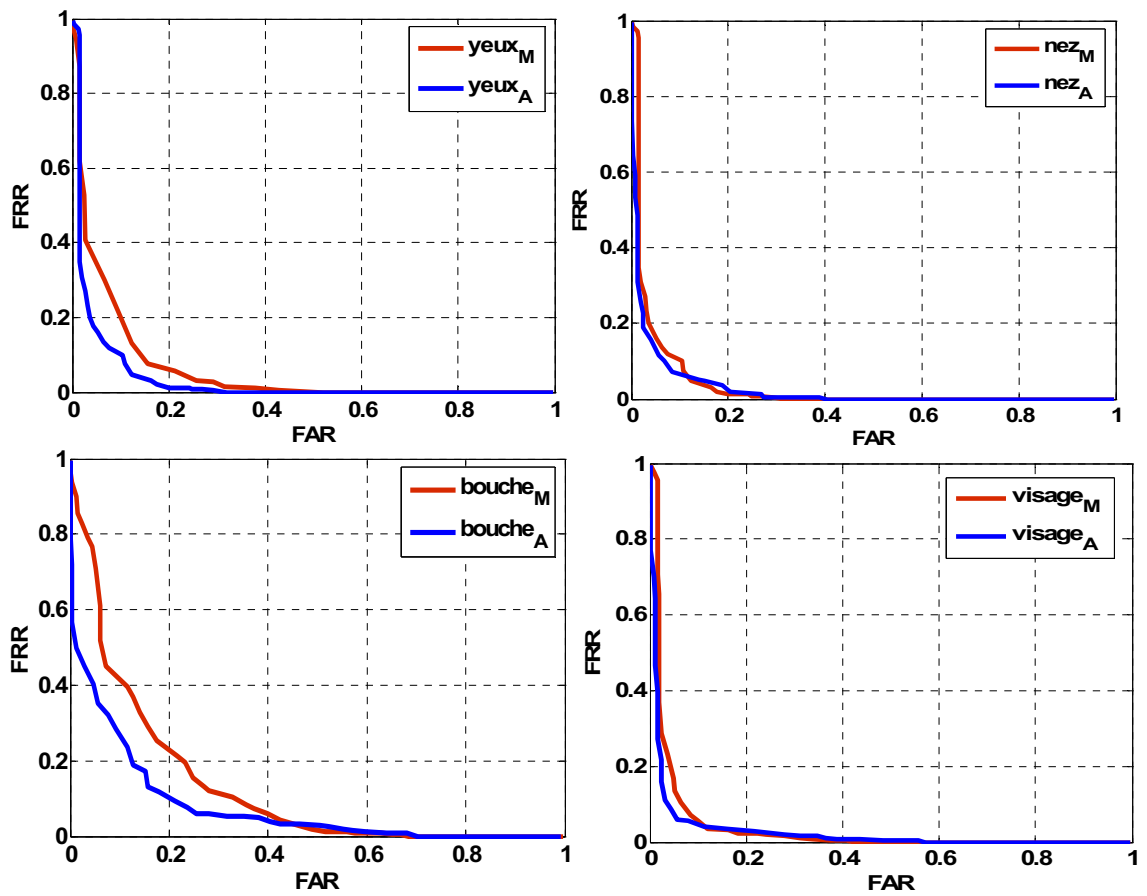


Figure 5. 7. Courbes DET : région_M extraction manuelle, région_A extraction automatique.

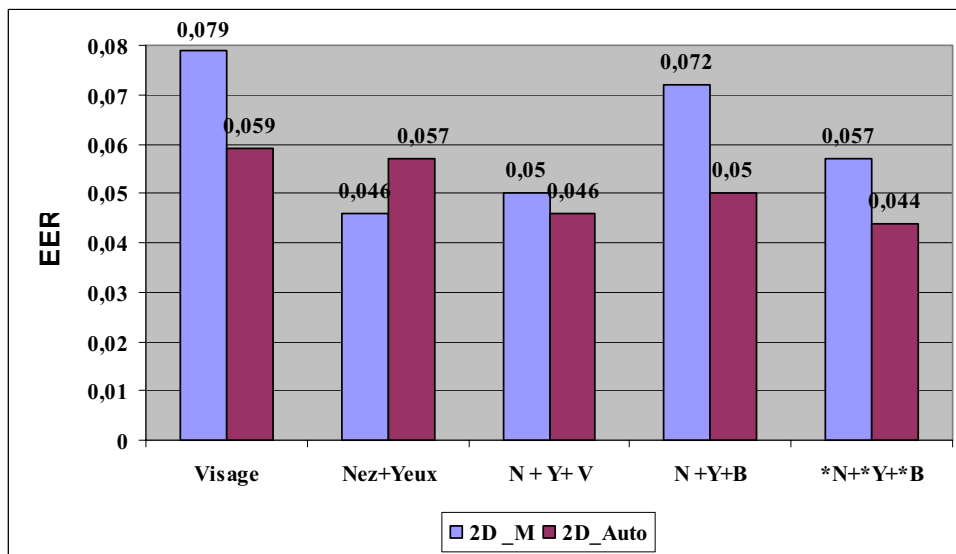


Figure 5. 8. Les différents résultats de fusion de classifieurs.

5.3 Vérification de visage 3D

Il existe plusieurs méthodes géométriques ou locales de reconnaissance 2D de visage (voir chapitre 2) robustes aux variations d'expression et aux occultations, mais pas vis-à-vis des changements de pose [Tan06]. Par ailleurs, peu de travaux ont été publiés sur des approches géométriques 3D, et comme nous l'avons souligné dans le chapitre 2, dans ces approches le choix des points caractéristiques n'est basé sur aucune étude de la structure du visage humain [Hus05][Wan02][Lee05]. Dans cette partie nous proposons une méthode géométrique de reconnaissance basée sur des mesures anthropologiques 3D de visage. L'utilisation de l'information 3D de visage est motivée par plusieurs raisons. Premièrement, le fait de travailler sur la forme purement géométrique du visage, plutôt que sur la couleur et la texture des images, rend le système de reconnaissance invariant à l'illumination. Deuxièmement, il est plus facile de changer l'orientation d'un modèle facial 3D dans l'espace pour compenser les variations de poses, qu'un visage sur une image 2D. Enfin, les informations de profondeur fournissent un atout supplémentaire pour la reconnaissance. Cela dit, il est évident que la couleur et la texture sont aussi une source riche d'information qui pourrait contribuer à améliorer les performances des systèmes de reconnaissance faciale. Ainsi, une combinaison des informations 3D et 2D du visage permettrait de réduire le taux d'erreur c'est que nous présentons dans la section 5.4.

Nous allons étudier dans cette partie, la faisabilité d'un système de reconnaissance du visage basé sur des données 3D seules, et comparer ses performances avec celles d'un système de reconnaissance 2D classique. Pour cela, nous commencerons par présenter une étude sur l'anthropométrie. Ensuite nous expliciterons le choix de la signature 3D retenue ainsi que l'extraction et la reconstruction des points 3D. Enfin nous exposerons les résultats des différents tests d'évaluation.

5.3.1 Etude de l'anthropométrie du visage

Le visage humain constitue la composante essentielle pour la reconnaissance d'une personne. Il est primordial alors d'étudier sa morphologie, son anatomie et son évaluation au cours de sa croissance et du vieillissement pour en tirer les points les plus discriminants et les plus stables. Les anthropologues ont mené des études sur les caractéristiques

métriques et descriptives du visage humain. Récemment, Bennaceur et al. [Ben05] ont fait une étude intéressante sur le développement du visage humain depuis la naissance jusqu'à l'âge adulte. Les auteurs ont résumé les mensurations du visage qu'ils ont appelées «données anthropométriques du vivant » et qui peuvent être utilisées pour l'identification de visage.

Deux méthodologies anthropométriques sont généralement utilisées [Ver04], la première adopte la comparaison des mesures réalisées entre différents points anatomiques ou construites sur les différentes photos. Pour assurer la validité de cette méthode, les photos utilisées doivent être préalablement normalisées à la même échelle. Cela est réalisé en fixant une distance de 6 cm entre les deux pupilles [Por00]. Les mesures sont ensuite prises en utilisant un compas numérique (digital calipers). Les résultats sont présentés dans une table récapitulative et pourront ainsi être comparés. La deuxième technique consiste à utiliser des indices et des angles calculés à partir de mesures effectuées entre des points précis, bien définis sur les différentes régions faciales et reconnus pour leur stabilité, permettant ainsi de travailler sans tenir compte de l'échelle des photos utilisées. Cette technique est utilisée pour mesurer les proportions cranio-facial. Ces proportions correspondent à des rapports de distances de paires de segments-de-droite entre des points spécifiques sur le crâne et des points faciaux [Far87]. Par exemple, l'indice nasal le plus utilisé est le rapport entre la largeur horizontale du nez et sa hauteur verticale. Les mesures Cranio-faciales sont utilisées dans différents domaines : dans la sculpture pour créer des visages idéaux bien proportionnés, dans l'anthropologie pour analyser des restes humains préhistoriques [Com60], et plus récemment en vision par ordinateur (estimation de l'orientation de la tête [Hor97], détection des points caractéristiques de visage [Soh06]) et en infographie [DeC98] pour créer des modèles paramétriques de visages humains. Dès 1939, Hrdlicka a souligné l'importance des mesures faciales anthropométriques pour comparer des groupes de personnes ou des populations [Hrd39]. Malgré cela, l'anthropométrie n'a pas été ou peu utilisée par la communauté de biométrie pour la reconnaissance 3D de visage. Après une étude approfondie de la littérature antérieure sur l'anthropométrie, Farkas et al. ont résumé une liste de 129 proportions anthropométriques cranio-faciales de base ayant été employées pour des tâches différentes [Far80]. Les auteurs ont aussi enregistré la moyenne et l'écart-type de ces indices en collectant des

mesures sur 1312 sujets humains adultes appartenant à des ethnies, sexes et tranches d'âge différents. Dans [Far87] les auteurs ont sélectionné 25 indices anthropométriques (voir l'annexe C). Ce choix fut motivé par le fait que les points repères devaient être identifiables sur des photographies.

Une des méthodes qui utilise les indices et les angles est la méthode de craniophotocomparaison, mise au point par [Per96]. Cette technique consiste à comparer un crâne avec une photographie (figure 5.9) d'une personne disparue. La comparaison prend en compte les valeurs indiciaires et angulaires permettant ainsi de travailler sans tenir compte de l'échelle des documents. Dans ce qui suit nous décrivons cette méthode sur un exemple concret, afin d'illustrer le choix et le calcul des différents paramètres utilisés pour la reconnaissance du visage. Sur la vue faciale du crâne et sur la photographie du sujet disparu (figure 5.9) sont positionnés des points anatomiques qui, reliés entre eux selon des critères précis, fournissent des paramètres (tableau 5.3), des indices et des valeurs angulaires. Il est important de noter que l'étude comparative ne prend jamais en compte la comparaison des valeurs brutes d'un même paramètre sur les deux clichés mais celle des rapports indiciaires confrontant les paramètres deux à deux, dans chaque cliché, ce qui offre l'avantage considérable de pouvoir travailler sur des instantanés n'étant pas à la même échelle! La ressemblance entre le crâne inconnu et la photographie va être établie en prenant en compte la différence algébrique des valeurs indiciaires (ou angulaires). Ensuite la somme algébrique (tableau 5.4) de l'ensemble des intervalles indiciaires est calculée puis divisée par le nombre d'indices pris en compte : le résultat ainsi obtenu (= moyenne algébrique) va permettre l'identification (c'est à dire l'assimilation crâne inconnu / visage), qui sera considérée comme : fortement probable (90%) à certaine (100%) dans le créneau : - 1 / 0 / + 1 .

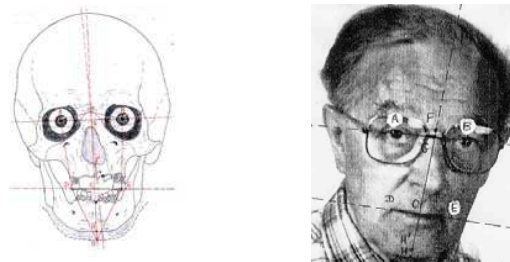


Figure 5. 9. Dessin au dioptrigraphe du Crâne1 montrant les points et paramètres utilisés dans la craniophotocomparaison

Points	Paramètres
A	Pupille droite
B	Pupille gauche
AB	Diamètre interpupillaire / plan du regard
DE	Ouverture de la bouche / plan occlusal
F	Nasion
F'	Nasospinal
F F'	Axe nasal
H	Gnathion*
H	Prognathion [point mentonnier]*
FH	Axe sagittal facial
C	Intersection FH / DE
G	Intersection FH / AB

Tableau 5. 3. Points et paramètres utilisés dans la craniophotocomparaison (d'après Desbois, Perrot et Desbois, 2004)

Résultats de la craniophotocomparaison crâne 1/ Mr C (Figure 5.9)

Angles	Crâne1	Mr C	Différence algébrique
FAD	97	87	10
GAD	82	85	-3
FBE	100	93	7
GBE	87	81	6
ADH	164	158	6
BEH	156	160	-4
Indices			
ABx100/DE	122	141.46	-19.46
FGx100/FC	10.48	10.81	-0.33
AGx100/FC	40.63	40.24	0.39
ABx100/FH	48.03	56.86	-8.83
AFx100/FC	41.94	40.95	0.99
Somme algébrique			-5.44
Moyenne algébrique			-0.49

Tableau 5. 4. Angles et valeurs indiciaires pris en compte dans la craniophotocomparaison [Des04]

Nous nous sommes inspirés de ces travaux pour définir les différents paramètres indiciaires et angulaires utilisés dans notre approche de reconnaissance 3D.

5.3.2 Sélection des points caractéristiques

Le choix des points caractéristiques du visage est très critique pour les méthodes de reconnaissance 3D. Ces points doivent être discriminants et stables afin de permettre l'identification de personnes sous différentes conditions d'éclairage, de poses et d'expressions faciales. Pour définir les points les plus pertinents du visage, nous nous sommes basés sur les résultats des recherches de la psychologie cognitive sur le système visuel humain. En effet, ces recherches ont clairement mis en évidence l'importance des régions des yeux, du nez et de la bouche pour l'identification faciale. Et donc les points caractéristiques utilisés doivent couvrir l'ensemble de ces trois régions. Par ailleurs, plusieurs recherches sur les méthodes géométriques et locales 2D ont montré que la partie supérieure du visage est la plus stable [Bru88][She81]. Les tests d'évaluation que nous avons effectués sur la base IV² et présentés dans la section précédente ont aussi démontré que la région de la bouche diminue considérablement le taux d'authentification surtout lorsque la base contient des changements significatifs d'expressions. Partant de ce constat, nous avons choisi les 12 points caractéristiques (voir figure 5.11-a) suivants :

Région yeux

À la différence de la méthode « craniocomparaison », nous avons choisi les 6 points les plus stables au lieu de deux points, car dans notre cas il s'agit de personnes vivantes :

- les quatre coins des yeux (P_1, P_3, P_4, P_6), deux centres pupilles (P_2, P_5).

Région nez

Pour qu'on puisse mesurer la largeur et la hauteur du nez nous avons choisi cinq points :

- le nasion (P_7), la pointe (P_8), le subnasal (P_{10}) et les deux extrémités du nez (P_9, P_{11}).

Région bouche

Nous avons choisi le point le plus stable en cas d'expression faciale, il s'agit du point (P_{12}) obtenu à partir de l'intersection des deux droites, L_1 reliant les deux cheïlons de la bouche et $L_2(P_7, P_{11})$. Il correspond au centre de la bouche.

La figure 5.10-a permet de visualiser ces 12 points.

5.3.3 La signature 3D

L'ensemble des points caractéristiques sélectionnés est utilisé pour calculer une signature qui doit permettre d'identifier sans ambiguïté le visage de la personne en question.

La signature 3D (voir tableau 5.6) que nous avons mise au point permet de caractériser une personne de façon unique. Elle correspond à un vecteur de distances 3D, d'indices et d'angles. Nous nous sommes basées sur plusieurs travaux de recherche [Ben05][Per96][Far87] pour choisir les paramètres géométriques utilisés dans le calcul de la signature 3D (voir tableau 5.5). Ainsi, selon Bennaceur [Ben05] les distances A_1 , A_2 , A_3 correspondent à des mensurations ayant une grande importance car elles permettent d'apprécier une partie de l'équilibre craniofacial. Les distances B_2 et B_1 sont choisies pour calculer le rapport entre le nez et la bouche [Far87]. Les distances B_2 et D sont choisies pour calculer l'indice du nez. Enfin, comme nous ne travaillons pas sur des crânes mais sur des visages de personnes vivantes, nous avons choisi de calculer les angles à partir des points les plus stables du visage (voir figure 5.12-c).

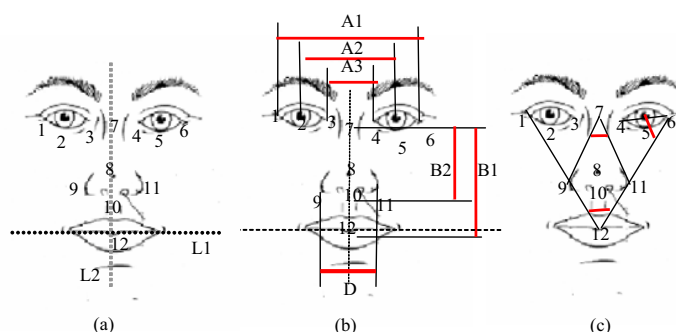


Figure 5. 10. Points (a), paramètres (b) et angles (c) utilisés pour la création d'une signature 3D facial.

Les tableaux suivants résument les éléments de la signature proposée :

Points	Paramètres
$A_1 = \text{dist}(1-6)$	Distance entre les coins extérieurs des yeux (la distance intercanthale externe)
$A_2 = \text{dist}(2-5)$	Distance interpupillaire
$A_3 = \text{dist}(3-4)$	Distance entre les coins intérieurs des yeux (la distance intercanthale interne)
$D = \text{dist}(9-11)$	Distance entre les deux extrémités du nez (la largeur du nez)
$B_1 = \text{dist}(7-12)$	Distance entre nasion et centre de la bouche
$B_2 = \text{dist}(7-10)$	Distance entre nasion et le subnasal ou hauteur du nez
$B_3 = \text{dist}(3-9)$	Distance entre le coin interne et l'extrémité du nez

Tableau 5. 5. Les paramètres utilisés pour la création de la signature 3D.

Signature	Paramètres
Distances	$A_1 = \text{dist}(P_1, P_6)$
	$A_2 = \text{dist}(P_2, P_5)$
	$A_3 = \text{dist}(P_3, P_4)$
	$B_1 = \text{dist}(P_7, P_{10})$
Indices	$100 * A_3 / A_1$
	$100 * B_2 / B_1$
	$100 * D / B_2$ (Indice du nez)
	$100 * A_2 / B_2$
Angles	$\text{Angl}(7-6-12)$
	$\text{Angl}(9-7-11)$
	$\text{Angl}(1-12-6)$

Tableau 5. 6. La signature 3D proposée

5.3.4 Système d'acquisition de visage 3D

L'inconvénient de la reconnaissance 3D du visage réside principalement dans le système d'acquisition. Les systèmes utilisés actuellement, comme le scanner laser, exigent une coopération active de l'utilisateur, ce dernier doit se tenir à une distance précise du scanner et rester immobile durant un certain laps de temps. De plus, le coût de ces systèmes reste très élevé, et le temps d'acquisition et du traitement des données 3D trop lent. Tous ces inconvénients ont été pratiquement constatés lors des acquisitions des visages 3D dans le cadre du projet Techno vision IV².

Dans le cadre de notre thèse, nous proposons d'utiliser la vision binoculaire ou vision stéréoscopique (utilisation de deux caméras fixes) pour générer les données 3D des visages. La vision stéréoscopique se veut la plus proche possible de la stratégie perceptive humaine, plaçant les caméras sur le même axe de vision figure (5.11), dans une configuration convergente. Chaque image droite et gauche représente la même scène vue sous un angle différent. La mise en correspondance entre les points des deux images permet, par triangulation, de reconstruire la scène 3D.



Figure 5. 11. Système stéréoscopique utilisé

Il est vrai que la vision stéréoscopique souffre d'un inconvénient majeur, à savoir le problème de l'appariement des points dense ou non dense, de primitives extraits des images gauche et droite. La précision des modèles reconstruits dépend essentiellement de la précision de la mise en correspondance. Ce problème est encore plus complexe dans le cas du visage. Dans notre cas, ce problème ne se pose pas, car nous n'avons pas besoin de reconstruire tout le visage mais uniquement quelques points caractéristiques (exemple : les coins des yeux).

5.3.4.1 Calibration des cameras

Avant de commencer les acquisitions, nous avons tout d'abord calibré les deux cameras. Plusieurs méthodes de calibration existent dans la littérature [Zho00] [Tsa87]. Nous avons utilisé la méthode de Toscani [Fau93] détaillée dans l'annexe B. Le calibrage consiste à calculer les paramètres intrinsèques et extrinsèques de la caméra à partir de points 3D dont on connaît la position dans un repère monde. Ce type d'objet aux coordonnées 3D connues est appelé mire de calibrage (voir figure 5.12). Le calibrage est alors réalisé pour chaque caméra indépendamment. On obtient deux matrices de projection perspective qui permettent de retrouver la géométrie euclidienne, lorsque le même point dans les deux images droite et gauche est mis en correspondance. Les données nécessaires sont : les coordonnées 3D de points dans l'espace monde et leurs correspondances 2D dans l'espace image. En pratique, plus 8 correspondances (point 3D/pixel 2D) sont nécessaires pour résoudre ce système d'équations et obtenir la matrice de projection affine. Dans notre cas nous avons utilisé 96 correspondances 2D/3D.

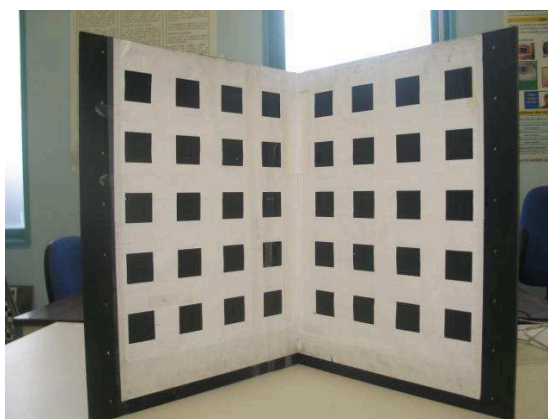


Figure 5. 12. Mire de calibration

5.3.4.2 Reconstruction des points 3D

La détection automatique des points caractéristiques du visage dans une image est un problème complexe qui fait l'objet de recherches intensives. Nous avons cité dans le chapitre 2 quelques travaux sur la reconnaissance 2D qui utilisent des méthodes automatiques d'extraction de points caractéristiques [Bru93] [Lee96] . L'objectif de cette thèse étant de proposer puis évaluer des méthodes de reconnaissance 2D/3D, nous n'avons

donc pas abordé le problème de la détection automatique des points caractéristiques, nous les avons extraits manuellement sur les images gauche et droite de la paire stéréoscopique.

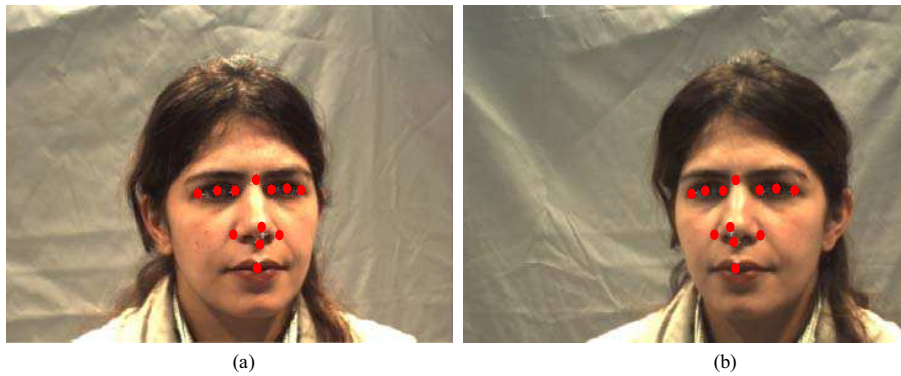


Figure 5.13. Différents points choisis pour la reconstruction.

La mise en correspondance consiste à trouver sur les deux images des paires de points homologues. Dans un premier temps, il s'agit de détecter les points caractéristiques dans chacune des images (voir figure 5.13), puis dans un second temps, les points de la première image sont appariés avec les points de la seconde image.

Dans notre cas la phase de mise en correspondance des points est faite automatiquement parce que les points caractéristiques sont toujours extraits dans le même ordre. Ces paires de points images appariés sont ensuite utilisées pour reconstruire les points 3D correspondants [Fau93].

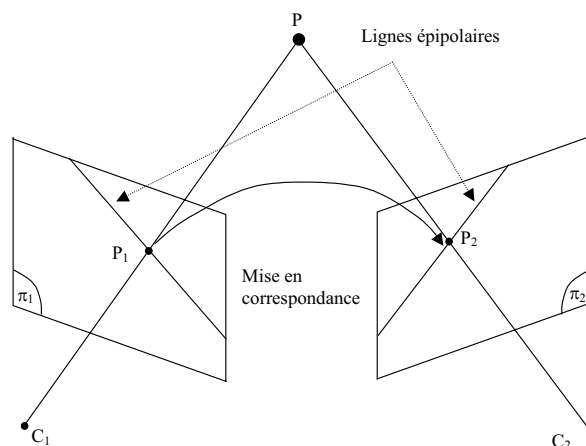


Figure 5.14. Vision stéréoscopique.

La connaissance des deux matrices de projection N , M des deux caméras est suffisante pour calculer les trois coordonnées d'un point quelconque P , connaissant ses images p_1 et p_2 (voir figure 5.14). En effet on peut écrire la transformation mire/image ainsi :

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} M_1 & M_{14} \\ M_2 & M_{24} \\ M_3 & M_{34} \end{pmatrix} \cdot \begin{pmatrix} P \\ 1 \end{pmatrix} \quad (5.17)$$

avec $(u, v)^t$ les coordonnées du point dans le repère image, s est un facteur d'échelle, $M_i=(M_{i1} \ M_{i2} \ M_{i3})$ et $P=(X \ Y \ Z)^t$. En développant l'équation (5.15) on aura le système suivant :

$$\begin{cases} su = M_1 P + M_{14} & (5.14) \\ sv = M_2 P + M_{24} & (5.15) \\ s = M_3 P + M_{34} & (5.16) \end{cases}$$

- en remplaçant (5.16) dans les deux équations (5.14) et (5.15), on obtient :

$$\begin{aligned} (M_1 - u_2 M_3)P &= u_2 M_{34} - M_{14} \\ (M_2 - v_2 M_3)P &= v_2 M_{34} - M_{24} \end{aligned} \quad (5.18)$$

Pour la deuxième caméra on aura un système identique et une autre matrice de projection N :

$$\begin{aligned} (N_1 - u_2 N_3)P &= u_2 N_{34} - N_{14} \\ (N_2 - v_2 N_3)P &= v_2 N_{34} - N_{24} \end{aligned} \quad (5.19)$$

Finalement, on aura un système à quatre équations et trois inconnues qui sont les coordonnées du point P :

$$\begin{cases} (M_1 - u_1 M_3)P = u_1 M_{34} - M_{14} \\ (M_2 - v_1 M_3)P = v_1 M_{34} - M_{24} \\ (N_1 - u_2 N_3)P = u_2 N_{34} - N_{14} \\ (N_2 - v_2 N_3)P = v_2 N_{34} - N_{24} \end{cases} \quad (5.20)$$

La solution de ce système est faite par les moindres carrés en résolvant l'équation : $AP=B$

$$\text{où : } A = \begin{pmatrix} M_1 - u_1 M_3 \\ M_2 - v_1 M_3 \\ N_1 - u_2 N_3 \\ N_2 - v_2 N_3 \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} u_1 M_{34} - M_{14} \\ v_1 M_{34} - M_{24} \\ u_2 N_{34} - N_{14} \\ v_2 N_{34} - N_{24} \end{pmatrix}$$

La solution aux moindres carrés est alors donnée par :

$$P = (A^t A)^{-1} A^t B \quad (5.21)$$

5.3.5 Protocole d'évaluation

Dans la littérature on parle souvent d'évaluation des techniques de reconnaissance de visages 3D sur des bases de visage 2.5D/3D. Afin d'évaluer notre approche sur une vraie base stéréoscopique nous avons utilisé les données de la base IV² car c'est la seule base en notre possession qui contient des images stéréoscopiques avec les métadonnées associées (notamment les matrices intrinsèque et extrinsèque des caméras), des informations essentielles pour la reconstruction 3D des points caractéristiques.

5.3.5.1 Extraction des données

A partir des vidéos de la base stéréo, nous avons extrait des images de 85 personnes. Pour chaque personne nous avons pris 4 paires d'images (4 pour l'image gauche et 4 pour l'image droite) sous des conditions différentes d'expression, d'illumination et de pose (voir figure 5.15).

Nous avons utilisé le logiciel VirtualDub pour extraire les images gauche et droite à partir des séquences vidéo. Il arrive que la synchronisation entre les deux caméras soit imparfaite. Nous avons dû extraire toutes les images de la vidéo et choisir les images gauche qui correspondent le mieux aux images droites. La résolution des images ainsi générées est moyenne (650x480).

Pour chaque paire d'images de personne nous réalisons les opérations suivantes :

- extraction manuelle des points caractéristiques 2D

- reconstruction des points 3D
- création de la signature 3D



Figure 5. 15. Images stéréo extraites de la base d'évaluation IV² : a) image droite, b) image gauche

5.3.5.2 Vérification 3D

La phase d'extraction nous a permis de créer 340 signatures (4 signatures par personne). Afin d'évaluer nos algorithmes en mode de vérification, nous avons utilisé une mesure de similarité qui correspond à une distance entre les vecteurs caractéristiques des images de test. Le nombre de comparaisons effectuées est égal à 510 dont 255 comparaisons intra classes. Comme pour l'approche 2D nous avons également choisi le taux EER pour mesurer le taux de vérification.

Par ailleurs, afin d'étudier la stabilité de la signature 3D nous proposons de calculer les écarts type intra-classe et inter-classe de chaque composante de la signature pour un ensemble de 85 personnes. Le tableau 5.7 montre un exemple de signatures 3D d'une même personne pour différentes expressions faciales.

Signature	Mr C neutre	Mr C sourir	Mr C surprise	Mr C dégoût
$A_1 = \text{dist}(P_1, P_6)$	90,167	90,88	91,608	88,157
$A_2 = \text{dist}(P_2, P_5)$	64,519	64,897	63,981	64,692
$A_3 = \text{dist}(P_3, P_4)$	37,987	38,536	36,913	38,335
$B_1 = \text{dist}(P_7, P_{10})$	50,028	50,217	50,71	50,303
$100 * A_3 / A_1$	122,41	120,11	124,69	125,34
$100 * B_2 / B_1$	45,325	46,003	44,393	42,793
$100 * D / B_2$	121,63	120,55	127,65	129,74
$100 * A_2 / B_2$	58,454	59,451	56,011	55,033
Angl(9-7-11)	43,395	44,76	44,525	49,393
Angl(4-6-12)	66,643	64,706	64,201	67,491
Angl(1-12-6)	47,972	48,17	46,741	45,093

Tableau 5.7. Exemple de signatures 3D d'une même personne pour différentes expressions faciales

Les résultats de cette étude statistique sont résumés dans le tableau 5.8. Nous avons ainsi déterminé, d'une part, l'écart type moyen inter classe de chaque mesure de la signature pour les 85 personnes, et d'autre part, l'écart type moyen intra classe d'une personne neutre. Pour l'étude intra classe, les écarts types faibles caractérisent les mesures les plus stables. Le tableau 5.8 montre que l'écart entre les yeux (mesure A1) et la longueur du nez (mesure B1) sont les mesures les plus stables. Par ailleurs, pour l'étude intra classe, nous remarquons que les écarts type sont élevés ce qui démontrent l'instabilité de la signature pour caractériser deux personnes différentes.

Tableau 5. 8 Etudes statistique entre les mesures de la signature

Signature	Ecart type intra_classe	Ecart type inter_classe
$A_1 = \text{dist}(P_1, P_6)$	1,45	3,44
$A_2 = \text{dist}(P_2, P_5)$	0,29	2,44
$A_3 = \text{dist}(P_3, P_4)$	0,65	3,11
$B_1 = \text{dist}(P_7, P_{10})$	0,20	6,12
$100 * A_3 / A_1$	1,78	9,99
$100 * B_2 / B_1$	1,55	2,66
$100 * D / B_2$	2,2	4,84
$100 * A_2 / B_2$	1,45	4,55
Angl(9-7-11)	1,23	2,34
Angl(4-6-12)	1,1	1,9
Angl(1-12-6)	1,02	2,8

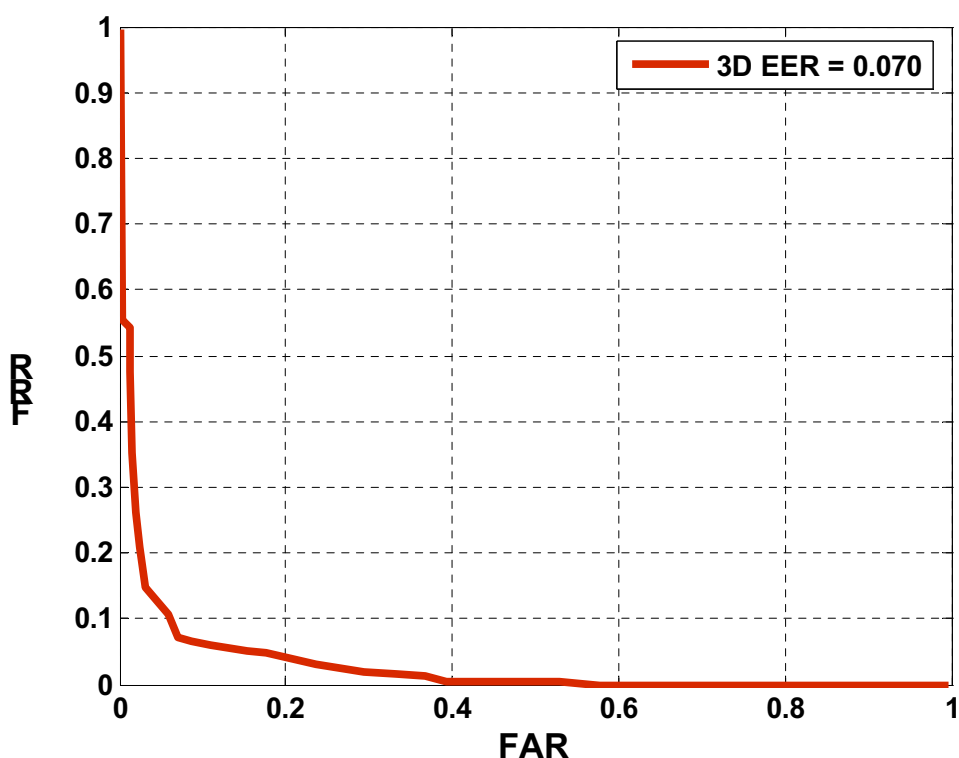


Figure 5. 16.Courbes DET : approches 3D.

5.3.5.3 Discussion

Les tests d'évaluation de notre approche de reconnaissance 3D de visage sur une sous base de IV^2 nous ont permis d'obtenir un EER très satisfaisant, de l'ordre de 0.070 ± 0.026 . Nous avons utilisé 510 comparaisons dont 255 comparaisons intra-classes. Par ailleurs, nous avons comparé nos résultats avec ceux fournis par les premiers tests d'évaluations de IV^2 sur des visages 3D (obtenus par scanner 3D). Ces tests utilisent 180 comparaisons de signatures (dont 60 comparaisons intra-classes) correspondant à des expressions différentes, pour évaluer deux algorithmes basés sur la méthode ICP, à savoir : R-ICP [Ben06] et ICP (implémenté par notre partenaire TALES). Le tableau (5.7) résume les taux EER réalisés par les trois approches. Nous remarquons que notre méthode (anthropométrie 3D) donne un EER satisfaisant ($EER=0.070 \pm 0.026$, figure 5.16) en comparaison avec les taux EER des deux autres approches. Toutefois, ce test présente une difficulté, liée au fait que l'on compare des approches différentes qui utilisent des capteurs d'acquisition et un nombre de signatures différentes. Néanmoins, il nous a permis de

mettre en évidence les performances de notre approche géométrique et de démontrer qu'elles étaient comparables aux performances des approches de reconnaissance du visage 3D.

Nb de comparaison	180 avec expression Intra-classe : 60		510 avec expression Intra-classe : 255
Approches	R-ICP	ICP	anthropométrie 3D
EER	0.067	0.083	0.070

Tableau 5. 7. Les taux EER obtenus par les différentes approches 3D.

Comme nous l'avons déjà mentionné dans les chapitres précédents, il existe de multiples facteurs qui peuvent entraîner une baisse du taux de vérification, notamment les variations de l'expression et de la pose. Dans notre cas, des facteurs supplémentaires dus à la nature même de notre approche peuvent s'ajouter et ainsi diminuer un peu plus le taux de vérification. En effet :

- la détection des points caractéristiques n'est pas évidente malgré le fait qu'elle soit réalisée manuellement, et ce à cause de la mauvaise qualité des images de la base et de la variation de l'illumination (visage ombré).
- les points 3D reconstruits par stéréovision sont moins précis que les points du visage 3D acquis avec un scanner laser.
- la base d'images stéréoscopiques sur laquelle nous avons effectuée nos tests contient un grand nombre d'images de visage présentant des variations importantes de pose et d'expression alors que la base de visage 3D est construite à partir des mesures du scanner laser sur des visages presque figés.

Dans la section suivante nous allons aborder l'approche multimodale qui combine les informations 2D et 3D du visage pour la reconnaissance faciale.

5.4 Vérifications de visage 2D/3D

Plusieurs études récentes ont montré que la fusion des deux modalités (2D et 3D) améliore nettement le taux de reconnaissance et d'authentification de visage [Bow06]. Il existe généralement deux façons pour combiner les modalités 2D et/ou 3D. La première consiste à combiner des données de même nature (2D ou 3D). Dans ce cas un seul type d'extracteur de signatures est utilisé. Dans notre approche modulaire, par exemple, l'extracteur employé est basé sur l'ACP. La deuxième consiste à utiliser des données de nature différente, et donc fait appel à des extracteurs de signatures différents.

Dans les sections précédentes, nous avons présenté en particulier deux approches pour la reconnaissance du visage (2D et 3D). La première est une approche 2D modulaire basée sur la fusion des scores, elle utilise une somme pondérée des scores de reconnaissance des trois régions du visage : yeux, nez et bouche. La deuxième est une approche 3D basée sur une signature anthropométrique 3D. Nous proposons ici de combiner ces deux approches afin de mettre en œuvre une nouvelle approche multimodale 2D/3D pour la reconnaissance du visage. La figure (5.17) montre un exemple de données utilisées pour la reconnaissance multimodale de visage. Il s'agit des trois régions : yeux, nez et bouche, et des 12 points 3D.

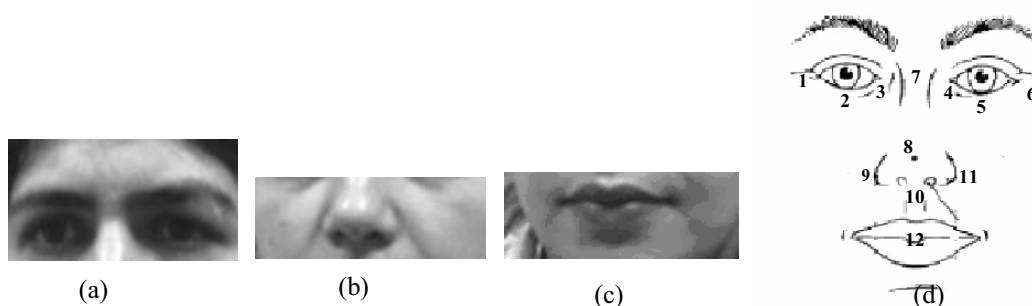


Figure 5. 17. Exemple des 4 données choisies pour la reconnaissance du visage

La figure (5.18) illustre le schéma de principe de la méthode multimodale 2D-3D. Le système commence par extraire les trois régions d'intérêt du visage (yeux, nez, bouche) à partir de l'image d'entrée droite. Ensuite il calcule un score pour chaque type de données

d'entrée. Le classifieur n° 1 travaille à partir de l'imagette contenant les yeux et les sourcils (de taille 50×65). Le classifieur n°2 travaille à partir de l'imagette du nez de taille 40×65, et enfin, le classifieur n° 3 travaille à partir de l'imagette de la bouche de taille 35×65. Par ailleurs, le système utilise les deux images d'entrée, droite et gauche afin de reconstruire les points d'intérêt 3D du visage lesquels sont ensuite utilisés pour calculer la signature 3D. Un 4^{ème} score est calculé à partir de cette signature 3D.

Pour les trois classifieurs 2D (yeux, nez et bouche) nous utilisons la somme pondérée (voir section 5.2.3.1), nous obtenons :

$$S_{2D} = 0.3112 * S_{Nez} + 0.3371 * S_{Yeux} + 0.3517 * S_{Bouche}$$

Le score S_{2D} est ensuite fusionné avec le score 3D en utilisant la méthode de la somme pondérée. Le résultat final de cette fusion donne le score de la méthode multimodale 2D-3D.

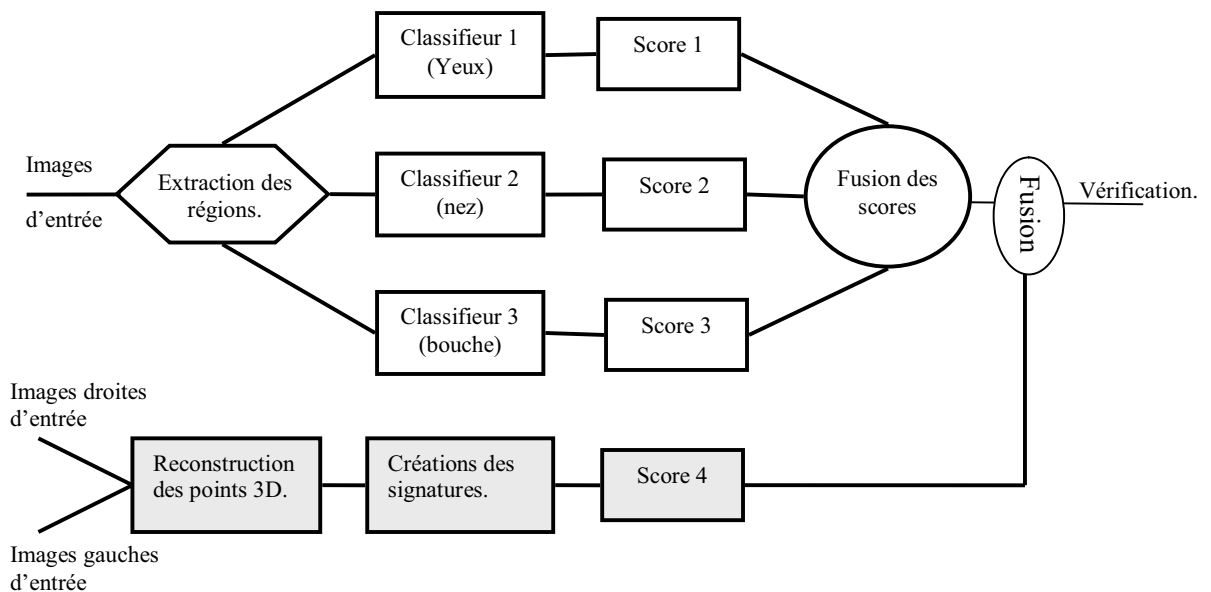


Figure 5. 18. Schéma de principe de l'approche multi modale 2D-3D

5.4.1 Evaluation de la méthode proposée

Nous avons pris pour chaque image 2D la signature 3D qui lui correspond. La sous-base utilisée dans ce cas est la même que celle utilisée pour la méthode de vérification 3D (c'est-à-dire 85 personnes, 4 images par personne). Nous avons tout normalisé les images

puis extrait les régions caractéristiques en suivant la procédure décrite dans la section (5.2.3.2).

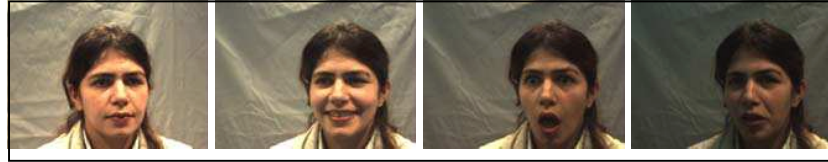


Figure 5. 19. Exemples d'images utilisées pour l'évaluation de l'approche 2D-3D.

Par ailleurs, pour construire les trois classifieurs 2D des trois régions (yeux, nez et bouche) nous avons pris au hasard 33 personnes (différentes des 85 personnes de la base de test), un exemple par personne pour la base d'apprentissage. Enfin, nous avons considéré les mêmes comparaisons que pour la méthode de vérification 3D, à savoir 510 comparaisons dont 255 intra classes.

5.4.2 Comparaison des approches 2D, 3D et 2D/ 3D

Nous avons effectué une première comparaison entre l'approche 2D modulaire et l'approche 3D. La figure (5.20) montre les courbes DET obtenues.

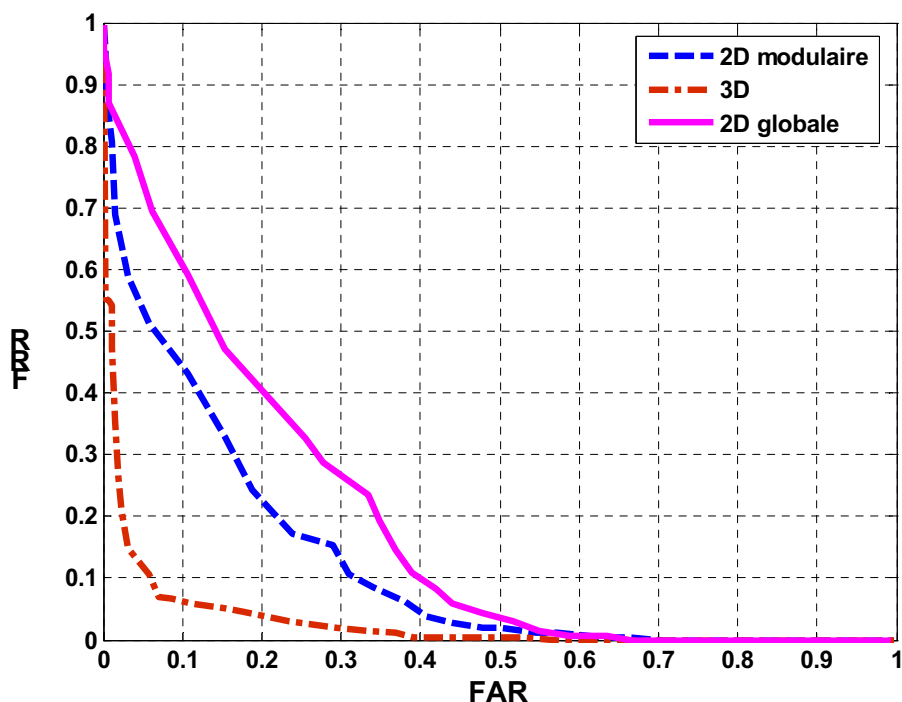


Figure 5. 20. Courbes DET : Comparaison entre les approches, 2D modulaire, 2D globale et 3D

Nous remarquons que l'approche 2D globale donne le plus mauvais taux $EER = 0.28 \pm 0.0490$ et le taux EER de l'approche modulaire égal à 0.21 ± 0.042 . Ce résultat nous a permis de définir une stratégie de fusion des scores 2D et 3D, en affectant un poids plus important au score 3D car celui-ci présente un EER plus petit ($EER_{3D} < EER_{2D}$) (voir figure 5.19). Ainsi :

$$S_{2D+3D} = \alpha_1 * S_{2D} + \alpha_2 * S_{3D}$$

Les valeurs des coefficients de pondération α_1 et α_2 sont déterminées en utilisant l'équation (5.14). Nous avons obtenu $\alpha_1 = 0.25$ et $\alpha_2 = 0.75$, et un taux $EER = 0.055 \pm 0.023$. Les taux obtenus pour les approches 2D et 3D seules sont $EER_{2D} = 0.21 \pm 0.042$ et $EER_{3D} = 0.070 \pm 0.026$ (voir figure 5.21).

Ce résultat illustre que la combinaison des informations 2D et 3D améliore nettement le taux de vérification, surtout lorsque la base de données utilisée contient des variations significatives de poses et d'expressions (ce qui est le cas de notre base).

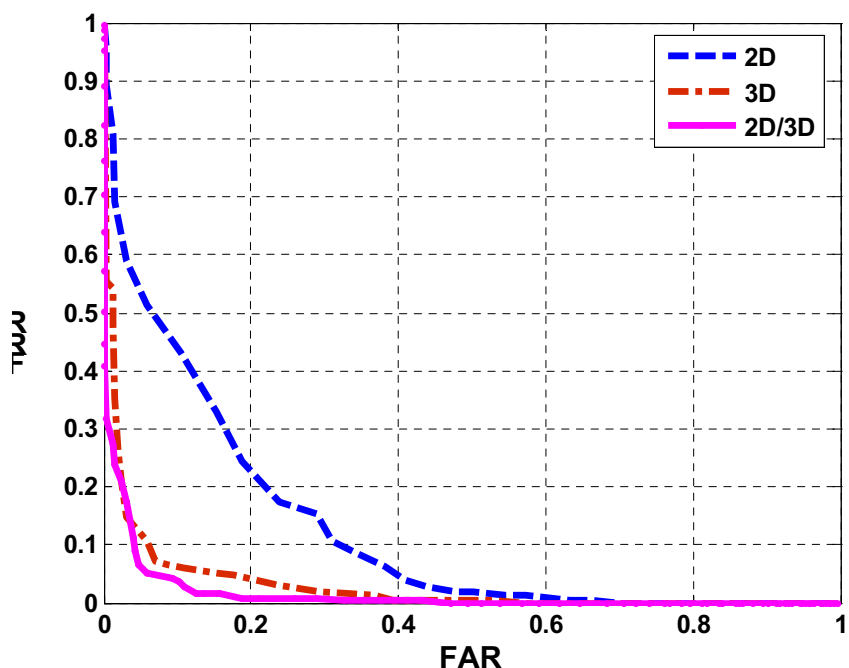


Figure 5. 21. Les Taux ERR des différentes approches.

5.4.3 Identification visage

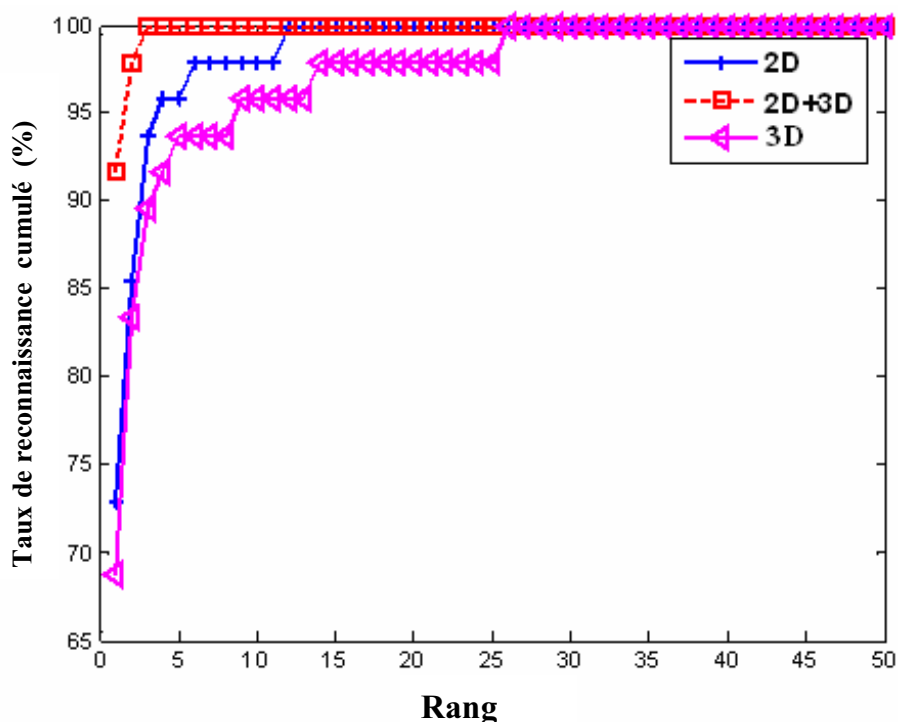
Pour compléter le protocole de test, nous proposons dans cette section d'évaluer nos différents algorithmes en mode d'identification. Pour cela, nous avons construit une matrice de similarité dans laquelle les signatures de test indexent les lignes et les signatures d'évaluation indexent les colonnes (ou inversement). Nous avons aussi comparé un autre mode d'agrégation, en l'occurrence le vote à la majorité (VM) [Kit98], avec la règle de la somme pondérée (SP). Enfin, nous avons comparé les taux de reconnaissance obtenus en utilisant le *score cumulé* (Cumulative Match Characteristics). En effet, il peut être utile de savoir si le bon choix se trouve dans les N premiers choix, notamment pour les applications d'aide à la décision.

5.4.3.1 Stratégie de la somme pondérée (SP)

Dans le tableau (5.7), nous avons donné, pour les trois approches, le taux de reconnaissance RR (Rank-one Recognition Rate) qui se limite au premier rang dans la liste triée des personnes. Nous constatons que l'approche 2D+3D donne le meilleur taux de reconnaissance et qu'à partir du Rang 3 ce taux atteint les 100 %.

Taux de reconnaissance	2D	3D	2D+3D
RR %	73 %	69 %	92%

Tableau 5. 8. Résultats de la reconnaissance : taux de reconnaissance au rang 1.

Figure 5. 22. Comparaison des courbes CMC de 2D, 3D et 2D/3D, sur la sous-base IV². Un visage est reconnu au rang r si une vue du même visage est parmi ses r plus proches voisins.

5.4.3.2 Stratégie Vote à la majorité (VM)

Le vote à la majorité consiste à assigner à l'image de test l'identité à laquelle elle est le plus fréquemment associée. En cas d'ambiguïté la signature 3D est le vainqueur. Le tableau (5.8) résume les résultats obtenus. Nous constatons que la somme pondérée (SP) donne le meilleur taux de reconnaissance RR (92%), ce qui nous conforte dans notre choix de la loi de fusion des scores.

Taux de reconnaissance	2D/3D - VM	2D/3D - SP
RR %	83 %	92%

Tableau 5. 9. Résultats de la reconnaissance : taux de reconnaissance au rang 1.

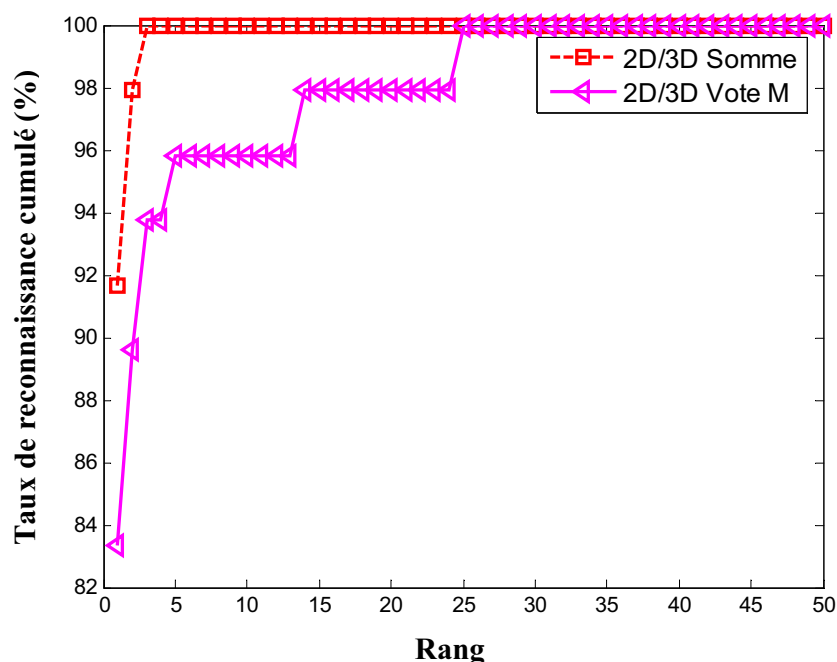


Figure 5. 23. Comparaison des courbes CMC 2D/3D (Vote à la majorité) et 2D/3D (Somme pondérée), sur la sous-base IV² phase 2.

5.5 Discussion et Conclusion

Dans ce chapitre, nous avons présenté trois approches de reconnaissance du visage, en l'occurrence une approche 2D modulaire, une approche 3D et une approche multimodale 2D/3D. Nous avons utilisé la base de données du projet IV² afin de tester et d'évaluer nos méthodes. L'approche ACP modulaire a été évaluée sur la base de données de la phase préliminaire de IV², nous avons constaté qu'elle donne les meilleures performances. En effet, l'approche modulaire permet de mieux gérer l'influence des conditions d'acquisition sur les différentes régions de visage.

Par ailleurs, nous avons démontré que, lorsque notre module de détection automatique de visage et de ses caractéristiques est utilisé au lieu d'une détection manuelle des régions, les performances de la reconnaissance 2D sont nettement améliorées.

Enfin, nous avons mis en évidence les performances nettement supérieures de l'approche multi-modale 2D-3D par rapport aux approches 2D modulaire et 3D. L'approche 2D+3D permet de réduire considérablement le taux de vérification ($EER = 0.05 \pm 0.023$) et d'augmenter le taux d'identification (92%).

Conclusion et Perspectives

Dans cette thèse, nous avons traité deux problématiques majeures et complémentaires rencontrées en reconnaissance de visage. Il s'agit, d'une part de l'extraction automatique de visage et de ses régions caractéristiques, et d'autre part de la reconnaissance du visage.

La difficulté de l'extraction de visage et de ses caractéristiques est due principalement aux variations des conditions d'éclairage. Nous avons donc proposé une nouvelle approche d'extraction qui s'adapte aux changements de l'illumination et de pose. Dans un premier temps, un algorithme efficace de segmentation couleur de visage dans l'espace TLS (Teinte, Luminance, Saturation) est appliqué sur l'image afin de déterminer les régions d'intérêt du visage. Pour cela, nous avons modifié la méthode de fusion des bassins versants obtenus à partir de l'algorithme de lignes de partage des eaux (LPE) en ajoutant un critère basé sur la pertinence de la teinte. Nous avons défini un domaine de la peau sur l'espace de la teinte afin d'extraire la région visage à partir de l'image segmentée. La méthode proposée pour l'extraction des zones du visage combine une classification basée sur la méthode de kmeans avec une approche géométrique afin d'identifier deux régions caractéristiques du visage, en l'occurrence les yeux et la bouche. Nous avons évalué notre approche sur deux bases de données : la base XM2VTS caractérisée par une grande variabilité d'apparence (lunettes, barbe, etc.), et la base IV² caractérisée par une variabilité de la lumière et de la pose.

Par ailleurs, nous avons pu travailler sur le problème de la reconnaissance de visage 2D/3D, dans le contexte de la vérification, grâce à la base de données multimodale IV². Nous avons ainsi évalué l'influence de la détection de régions sur la reconnaissance de visage. Pour cela, nous avons d'abord extrait les régions du visage manuellement puis de manière automatique

en utilisant notre approche d'extraction, et nous avons comparé les résultats de reconnaissance obtenus sur ces régions. Nous avons appliqué les méthodes classiques « EigenFace » et « EigenFace » modulaire et avons étudié différentes combinaisons pour la fusion des scores de reconnaissance. Les résultats obtenus ont montré que l'approche modulaire donne les meilleures performances, nous l'avons donc utilisée dans la suite de notre travail.

Lorsqu'on parle d'informations 3D, on sous-entend des mesures réelles effectuées sur la structure anatomique du visage : c'est l'anthropométrie faciale, une discipline consacrée à l'étude des proportions du visage humain à partir de son observation anatomique et physiologique. Nous avons proposé une approche locale 3D de reconnaissance de visage basée sur des mesures anthropométriques correspondants à des distances, des indices et des angles. Nous avons utilisé un système stéréoscopique qui repose sur deux camera fixes pour construire les points d'intérêt 3D de visage.

Enfin, nous avons proposé une approche multimodale 2D/3D qui repose sur la fusion pondérée des scores de l'approche « EigenFace » modulaire avec la signature anthropométriques 3D de visage. Nous avons évalué nos approches de reconnaissance 3D et 2D/3D sur une sous-bases de IV^2 qui contient des images stéréoscopiques de visage. Nous avons obtenus des résultats très intéressants en comparaison avec les techniques classiques de reconnaissance 2D de visage.

Les perspectives de ce travail sont nombreuses : dans un premier temps, nous souhaitons améliorer le module d'extraction automatique de visage et de ses caractéristiques. Actuellement, nous avons défini un domaine fixe de couleur de la peau dans l'espace teinte. Or, comme la teinte est liée à la saturation qui elle-même dépend de l'illumination, nous pensons que choisir un domaine variable pour la couleur de la peau améliorerait la robustesse de la méthode de détection vis-à-vis des changements d'illumination. Dans un deuxième temps, nous allons travailler sur l'automatisation des approches de reconnaissance élaborées. Pour cela nous proposons de développer des techniques automatiques de détection de points d'intérêt. Des techniques de type « template matching » ou basées snakes mériteraient d'être explorées. Enfin, Une évaluation plus poussée permettrait une meilleure validation des approches développées. Pour cela, nous allons utiliser la totalité de la base IV^2 et allons comparer les résultats de la reconnaissance basée stéréovision avec la reconnaissance basée sur les données 3D fournies par le scanner laser.

Annexe A

Algorithmes

Dans cette annexe nous présentons les différents algorithmes proposés :

Algorithme: Tri des pixels par norme du gradient croissants

- Entrée : I , image numérique ;
- Sortie : HC_I histogramme cumulé,
- I_I tableau trié de (pointeur de) pixels défini sur D_I
- Initialisations :
 - h_{min}, h_{max} : extrema de la norme du gradient dans I ;
 - H_I : tableau à valeurs entières indexées sur $[h_{min}, h_{max}]$;
 - $\forall h \in [h_{min}, h_{max}], H_I[h] \leftarrow 0$;
 - $HC_I[h_{min}] \leftarrow 0$;
- Pour tout pixel $p \in D_I$ { /*Histogramme*/
 - $H_I[I(p)] \leftarrow H_I[I(p)] + 1$; }
- Pour $h = h_{min} + 1$ à h_{max} { /* Histogramme cumulé */
 - $HC_I[h] \leftarrow HC_I[h-1] + H_I[h-1]$; }

Les structures de données utilisées sont les mêmes que dans [Vin91].

- Une image d'étiquettes
- Une file d'attente FIFO qui stocke les dilatations successives

Algorithme : Ligne de Partage des Eaux

```
# define MASK      -2 /* valeur initiale affectée à chaque niveau */
# define WSHED     0 /* valeur finale des pixels de la LPE */
# define INQUEUE  -3 /* valeur d'un pixel lors de son insertion dans la file*/
```

- - Entrée: I, image numérique ;
- Sortie: J, image des bassins versants étiquetés ;
label : 1,2,3, etc. Les pixels de la LPE sont mis à WSHED = 0 ;
- Initialisations :
 - $\forall p \in D_I, J(p) \leftarrow -1;$
 - label_courant $\leftarrow 0;$
 - flag : variable booléenne
- Fait appel à la fonction Tri

Soit h_{\min} et h_{\max} les valeurs extrémales de I;

- Pour $h = h_{\min}$ à h_{\max} {
 - /* SKIZ géodésique du niveau $h - 1$ dans le niveau h */
 - Pour tout pixel p tel que $I(p) = h$ {
 - /* Ces pixels sont accessibles directement grâce au tri initial*/
 - $J(p) \leftarrow MASK;$
 - Si $\exists p' \in N_G(p)$ tel que $J(p') > 0$ ou $J(p') = WSHED$ {
 - $J(p) \leftarrow INQUEUE; \text{fifo_ajouter}(p);$ }
 - }
 - Tant que $\text{fifo_vide}() = \text{faux}$ {
 - $p \leftarrow \text{fifo_retirer}();$
 - Pour tout pixel $p \in N_G(p)$ {

```

Si  $J(p) > 0$  {
    /*  $p$  appartient à un bassin déjà étiqueté */
    Si ( $J(p) = INQUEUE$  ou ( $J(p) = WSHED$  et  $flag = vrai$ ))
         $J(p) \leftarrow J(p)$ ;
    sinon si ( $J(p) > 0$  et  $J(p) \neq J(p')$ ) {
         $J(p) \leftarrow WSHED$ ;
         $flag \leftarrow faux$ ;
    }
}
sinon si  $J(p) = WSHED$  {
    si  $J(p) = INQUEUE$  {
         $J(p) \leftarrow WSHED$ ;
         $flag \leftarrow vrai$ ;
    }
}
sinon si  $J(p) = MASK$  {
     $J(p) \leftarrow INQUEUE$ ;
     $fifo\_ajouter(p)$ ;
}
}
}
/* De nouveaux minima sont-ils apparus ? */
- Pour tout pixel  $p$  tel que  $I(p) = h$  {
    Si  $J(p) = MASK$  {
         $label\_courant \leftarrow label\_courant + 1$ ;
         $fifo\_add(p)$ ;
         $J(p) \leftarrow label\_courant$ ;
        Tant que  $fifo\_vide() = faux$  {
             $p \leftarrow fifo\_retirer()$ ;
            pour tout pixel  $p' \in N_c(p)$  {
                Si  $J(p') = MASK$  {  $fifo\_add(p')$ ;  $J(p') \leftarrow label\_courant$ ; }
            }
        }
    }
}
}
}
}

```

Cette version de l'algorithme LPE utilise la variable flag afin d'obtenir une ligne de partage des eaux exacte. En effet, si la variable flag n'avait pas été disponible on peut avoir un problème d'épaississement de la LPE, comme illustré dans la figure A.1.

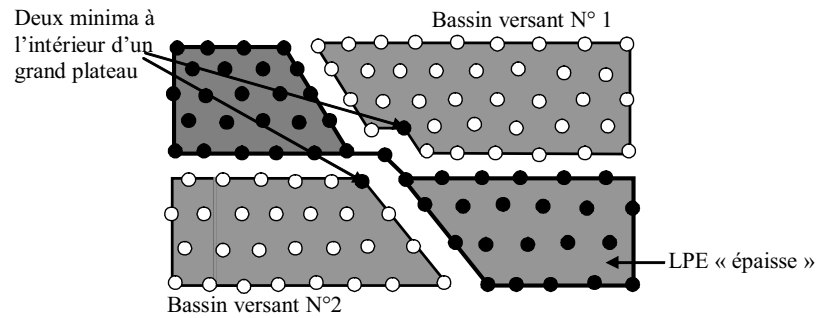


Figure A.1. Si l'on affectait systématiquement la valeur WSHEd à tout pixel ayant un pixel à la valeur WSHEd dans son voisinage, on pourrait observer des lignes de partage des eaux « épaisses ». La dissymétrie des bassins versants tracés ci dessus provient du fait que les voisins du minimum situé en haut à droite ont été placés dans la file d'attente avant ceux de l'autre minimum.

Annexe B

Calibration de la Caméra

Dans cette annexe nous allons décrire le modèle géométrique associé au processus de saisie d'images à l'aide d'une caméra [Hor95]. Ce modèle est caractérisé par un certain nombre de paramètres que nous allons estimer par calibration. Le modèle le plus utilisé pour modéliser une caméra est le modèle sténopé, ce dernier modélise une projection perspective avec un centre de projection F et un plan image (cf. figure B.1).

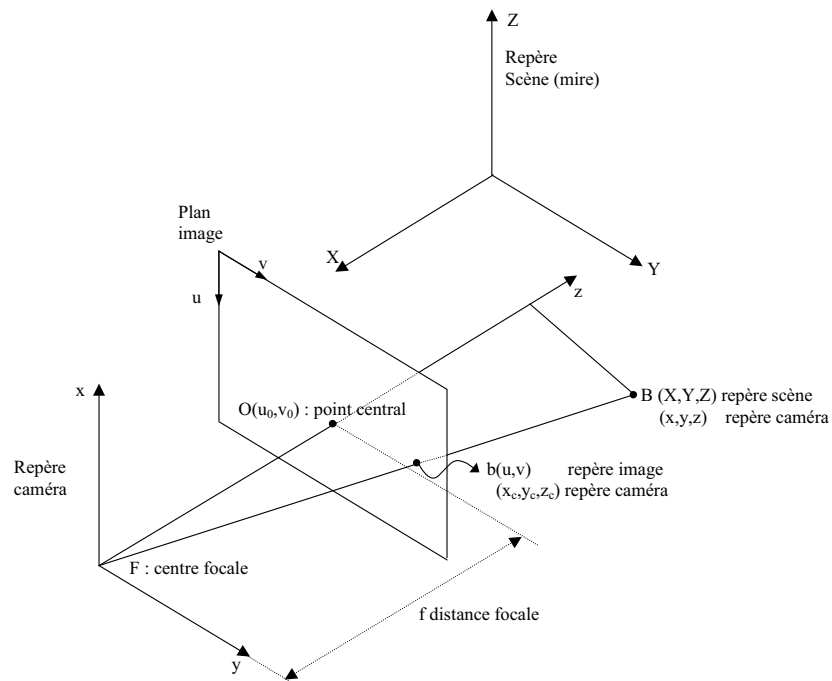


Figure B.1 : le modèle géométrique d'une caméra

Le modèle géométrique de la caméra est donc caractérisé par deux transformations:

1. une projection perspective qui transforme un point 3D en un point image 2D.
2. une transformation rigide du repère lié à la caméra au repère lié à l'image.

B.1 La projection perspective

Un point B de l'espace se projette dans le plan image de la caméra suivant une droite passant par B et le centre focale (centre de projection) F (cf. figure B.1).

Considérons un repère centré sur le centre de projection F et dont l'axe z soit orthogonal au plan image qu'on appellera repère caméra. Soient (x,y,z) les coordonnées du point B dans le repère caméra et (x_c, y_c, z_c) les coordonnées, dans le même repère, du projeté de B dans le plan image.

La projection perspective peut se modéliser sous la forme matricielle :

$$\begin{pmatrix} sx_c \\ sy_c \\ sz_c \\ s \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{z} & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (\text{B.1})$$

P

B.2 Transformation caméra/image

Les points image sont mesurés en pixels. Afin de pouvoir écrire la matrice de transformation du repère caméra au repère image, nous introduisons les paramètres suivants : (u_0, v_0) les coordonnées de F (en pixels) dans le repère image, k_u le facteur d'échelle vertical (pixels/mm) et k_v le facteur d'échelle horizontal. La transformation du repère caméra au repère image s'écrit alors (pour le point b) :

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} -k_u & 0 & 0 & u_0 \\ 0 & k_v & 0 & v_0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix} \quad (\text{B.2})$$

K

B.3 Les paramètres intrinsèques

En multipliant les matrices K et P (projection perspective suivie d'une transformation affine) nous pouvons écrire les équations du modèle géométrique de la caméra, soit la

relation entre les coordonnées caméra (x,y,z) du point B et les coordonnées image (u,v) du point b :

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (\text{B.3})$$

Ce modèle comporte quatre paramètres : $\alpha_u = -k_u \cdot f$, $\alpha_v = k_v \cdot f$, u_0 et v_0 . Ce sont les paramètres intrinsèques de la caméra, ils sont associés à la nature du capteur utilisé : sa focale f , la taille de la matrice du capteur et la géométrie des capteurs élémentaires qui forment la matrice de photo-transistors (k_u, k_v) . Ce sont ses paramètres qui vont être estimés par calibration.

B.3 Les paramètres extrinsèques

Afin de déterminer les paramètres du modèle de la caméra, nous allons placer devant la caméra une mire (un objet étalon) : un ensemble de points dont les coordonnées sont parfaitement connues dans un repère de la mire qui est différent du repère caméra (cf. figure B.1). Chaque point de la mire se projette dans l'image et on mesure ses coordonnées dans le repère image. La transformation mire/image se décompose donc en une transformation mire/caméra suivie d'une projection et suivie enfin d'une transformation caméra/image. La transformation mire/caméra se compose d'une rotation et d'une translation :

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \quad (\text{B.4})$$

La transformation rigide (rotation et translation peut s'écrire sous la forme d'une transformation homogène :

$$A = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix} \quad (\text{B.5})$$

Cette matrice représente les paramètres extrinsèques de la caméra. Ce sont les paramètres du déplacement rigide entre le repère de la mire et le repère caméra.

B.4 La transformation mire/image

Nous pouvons maintenant écrire la transformation mire/image sous la forme d'une matrice (3×4) appelée matrice de projection perspective et qui peut se décomposer comme suit :

$$M = I_c \cdot A = \begin{pmatrix} \alpha_u r_1 + u_0 t_x & \alpha_u t_x + u_0 t_z \\ \alpha_v r_2 + v_0 t_y & \alpha_v t_y + v_0 t_z \\ r_3 & t_z \end{pmatrix} \quad (\text{B.6})$$

Dans cette formule la matrice A a été écrite sous une forme compacte en utilisant la notation $\mathbf{r}_i = (r_{i1} \ r_{i2} \ r_{i3})$:

$$A = \begin{pmatrix} \mathbf{r}_1 & t_x \\ \mathbf{r}_2 & t_y \\ \mathbf{r}_3 & t_z \\ 0 & 1 \end{pmatrix} \quad (\text{B.7})$$

M est la matrice de projection perspective et elle peut, en général, s'écrire sous la forme suivante :

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{32} & m_{34} \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (\text{B.8})$$

Dans cette formule (X,Y,Z) sont les coordonnées du point B dans le repère de la mire. Cette matrice peut également s'écrire sous une forme plus simple :

$$M = \begin{pmatrix} \mathbf{m}_1 & m_{14} \\ \mathbf{m}_2 & m_{24} \\ \mathbf{m}_3 & m_{34} \end{pmatrix} \quad (\text{B.9})$$

avec $\mathbf{m}_i = (m_{i1} \ m_{i2} \ m_{i3})$. En identifiant M avec $I_c A$, en tenant compte des propriétés d'orthonormalité de la rotation, on obtient un ensemble d'équations qui permettent de calculer les paramètres intrinsèques et extrinsèques en fonction des coefficients de M. On obtient :

$$\begin{cases} \mathbf{r}_3 = \mathbf{m}_3 \\ u_0 = \mathbf{m}_1 \cdot \mathbf{m}_3 \\ v_0 = \mathbf{m}_2 \cdot \mathbf{m}_3 \\ \alpha_u = -\|\mathbf{m}_1 \times \mathbf{m}_3\| \\ \alpha_v = \|\mathbf{m}_2 \times \mathbf{m}_3\| \\ \mathbf{r}_1 = \frac{1}{\alpha_u}(\mathbf{m}_1 - u_0 \mathbf{m}_3) \\ \mathbf{r}_2 = \frac{1}{\alpha_v}(\mathbf{m}_2 - v_0 \mathbf{m}_3) \\ t_x = \frac{1}{\alpha_u}(m_{14} - u_0 m_{34}) \\ t_y = \frac{1}{\alpha_v}(m_{24} - v_0 m_{34}) \\ t_z = m_{34} \end{cases} \quad (\text{B.10})$$

Afin de trouver les paramètres intrinsèques et extrinsèques de la caméra on doit donc :

1. estimer les coefficients de la matrice de projection M et
2. extraire les paramètres de la caméra à partir de ces coefficients grâce aux formules données par l'équation (B.10)

B.5 Estimation des coefficients de la matrice M

En utilisant l'équation (A.8) on peut écrire les coordonnées image d'un point de la scène.

On obtient :

$$u = \frac{m_{11}X + m_{12}Y + m_{13}Z + m_{14}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \quad (\text{B.11})$$

$$v = \frac{m_{21}X + m_{22}Y + m_{23}Z + m_{24}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \quad (\text{B.12})$$

La matrice de projection M comporte 12 coefficients, or chaque points 3D de la mire fournit 2 équations (une pour u et l'autre pour v). Il nous faut donc au minimum 6 points distincts pour estimer les coefficients de la matrice de projection M. Ainsi pour n points on obtient 2n équations issues du réarrangement des équations (B.11) et (B.12) sous forme de combinaisons linéaires des m_{ij} , qu'on peut réécrire sous forme matricielle :

$$\mathbf{x}_9 = (m_1 \quad m_{14} \quad m_2 \quad m_{24} \quad m_{34})^T \quad \text{et} \quad \mathbf{x}_3 = (m_3)^T$$

On cherche donc à minimiser le critère Q :

$$Q = \|\mathbf{B}_{2n \times 9} \mathbf{x}_9 + \mathbf{C}_{2n \times 3} \mathbf{x}_3\|^2 \quad (\text{B.17})$$

avec la contrainte supplémentaire de l'équation (B.15), on peut réécrire (B.17) sous la forme:

$$Q = \|\mathbf{B} \mathbf{x}_9 + \mathbf{C} \mathbf{x}_3\|^2 + \lambda (1 - \|\mathbf{x}_3\|^2) \quad (\text{B.18})$$

Minimiser Q revient à déterminer les zéros de ses dérivées suivant \mathbf{x}_3 et \mathbf{x}_9 , on obtient alors deux équations:

$$\begin{cases} \frac{\delta Q}{\delta \mathbf{x}_9} = \mathbf{B}' \mathbf{B} \mathbf{x}_9 + \mathbf{B}' \mathbf{C} \mathbf{x}_3 = 0 \\ \frac{\delta Q}{\delta \mathbf{x}_3} = \mathbf{C}' \mathbf{C} \mathbf{x}_3 + \mathbf{C}' \mathbf{B} \mathbf{x}_9 - \lambda \mathbf{x}_3 = 0 \end{cases} \quad (\text{B.19})$$

d'où on obtient :

$$\begin{cases} \mathbf{x}_9 = -(\mathbf{B}' \mathbf{B})^{-1} \mathbf{B}' \mathbf{C} \mathbf{x}_3 \\ \mathbf{D} \mathbf{x}_3 = \lambda \mathbf{x}_3 \\ \mathbf{D} = \mathbf{C}' \mathbf{C} - \mathbf{C}' \mathbf{B} (\mathbf{B}' \mathbf{B})^{-1} \mathbf{B}' \mathbf{C} \end{cases} \quad (\text{B.20})$$

D est une matrice symétrique définie positive, elle a donc des valeurs propres réelles et positives. \mathbf{x}_3 est déterminé par le vecteur propre associé à la plus petite valeur propre λ de D.

La détermination des paramètres m_{ij} doit donc suivre l'algorithme suivant :

1. Calculer les valeurs propres de D.
2. Choisir la plus petite des valeurs propres λ pour minimiser le critère Q.
3. Obtenir le vecteur propre qui lui est associé, puis le normaliser pour ainsi obtenir \mathbf{x}_3 avec $\|\mathbf{x}_3\|=1$.
4. Puis calculer \mathbf{x}_9 à partir de la première équation de (B.20).

La matrice M est donc composée des valeurs extraites de \mathbf{x}_3 et \mathbf{x}_9 . Etant donné que le signe de \mathbf{x}_3 n'est pas défini, on a deux solutions, M et $-M$. On peut en choisir une parmi ces solutions en utilisant le fait que l'objet de calibration se trouve devant la caméra et non pas derrière la caméra. Dans ce cas on doit avoir $m_{34} = t_z > 0$.

Annexe C

Indices anthropométriques

Le tableau C.1 présente les différentes proportions utilisées par Farkas [Eld06]

Index number	Name of index	Description
1	Upper face-face height index	<u>Nasion-stomion</u>
2	Lower face-face height index	<u>Nasion-gnathion</u> <u>Subnasale-gnathion</u>
3	Mandibulo-face height index	<u>Nasion-gnathion</u> <u>Stomion-gnathion</u>
4	Mandibulo-upper face height index	<u>Nasion-gnathion</u> <u>Stomion-gnathion</u>
5	Mandibulo-lower face height index	<u>Nasion-stomion</u> <u>Stomion-gnathion</u> <u>Subnasale-gnathion</u>
6	Upper-middle third face depth index	<u>Tragion (l)-nasion</u> <u>Tragion (l)-subnasale</u>
7	Middle-lower third face depth index	<u>Tragion (l)-subnasale</u> <u>Tragion (l)-gnathion</u>
8	Nasal index	<u>Alare (r)-alare(l)</u> <u>Nasion-subnasale</u>
9	Upper lip height-mouth width index	<u>Subnasale-stomion</u> <u>Chelion (r)-chelion (l)</u>
10	Cutaneous-total upper lip height index	<u>Subnasale-labiale superius</u> <u>Subnasale-stomion</u>
11	Vermilion-total upper lip height index	<u>Labiale superius-stomion</u> <u>Subnasale-stomion</u>
12	Vermilion-cutaneous upper lip height index	<u>Labiale superius-stomion</u> <u>Subnasale-labiale superius</u>
13	Upper lip vertical contour index	<u>Subnasale-stomion</u> <u>Subnasale-labiale superius plus</u> <u>Labiale superius-stomion</u>
14	Vermilion height index	<u>Labiale superius-stomion</u> <u>Stomion-labiale inferius</u>
15	Chin-mandible height index	<u>Sublabiale-gnathion</u> <u>Stomion-gnathion</u>
16	Upper face height-biocular width index	<u>Nasion-stomion</u> <u>Exocanthion (r)-exocanthion (l)</u>
17	Intercanthal-nasal width index	<u>Endocanthion (r)-endocanthion(l)</u> <u>Alare (r)-Alare (l)</u>
18	Nose-face height index	<u>Nasion-subnasale</u> <u>Nasion-gnathion</u>
19	Nose-mouth width index	<u>Alare (r)-alare (l)</u> <u>Chelion (r)-chelion (l)</u>
20	Upper lip-upper face height index	<u>Subnasale-stomion</u> <u>Nasion-stomion</u>
21	Upper lip-mandible height index	<u>Subnasale-stomion</u> <u>Stomion-gnathion</u>
22	Upper lip-nose height index	<u>Subnasale-stomion</u> <u>Nasion-subnasale</u>
23	Lower lip-face height index	<u>Stomion-sublabiale</u> <u>Subnasale-gnathion</u>
24	Lower lip-mandible height index	<u>Stomion-sublabiale</u> <u>Stomion-gnathion</u>
25	Lower lip-chin height index	<u>Stomion-sublabiale</u> <u>Sublabiale-gnathion</u>

Annexe C

Indices anthropométriques

Le tableau C.1 présente les différentes proportions utilisées par Farkas [Eld06]

Index number	Name of index	Description
1	Upper face-face height index	<u>Nasion-stomion</u>
2	Lower face-face height index	<u>Nasion-gnathion</u> <u>Subnasale-gnathion</u>
3	Mandibulo-face height index	<u>Nasion-gnathion</u> <u>Stomion-gnathion</u>
4	Mandibulo-upper face height index	<u>Nasion-gnathion</u> <u>Stomion-gnathion</u>
5	Mandibulo-lower face height index	<u>Nasion-stomion</u> <u>Stomion-gnathion</u> <u>Subnasale-gnathion</u>
6	Upper-middle third face depth index	<u>Tragion (l)-nasion</u> <u>Tragion (l)-subnasale</u>
7	Middle-lower third face depth index	<u>Tragion (l)-subnasale</u> <u>Tragion (l)-gnathion</u>
8	Nasal index	<u>Alare (r)-alare(l)</u> <u>Nasion-subnasale</u>
9	Upper lip height-mouth width index	<u>Subnasale-stomion</u> <u>Chelion (r)-chelion (l)</u>
10	Cutaneous-total upper lip height index	<u>Subnasale-labiale superius</u> <u>Subnasale-stomion</u>
11	Vermilion-total upper lip height index	<u>Labiale superius-stomion</u> <u>Subnasale-stomion</u>
12	Vermilion-cutaneous upper lip height index	<u>Labiale superius-stomion</u> <u>Subnasale-labiale superius</u>
13	Upper lip vertical contour index	<u>Subnasale-stomion</u> <u>Subnasale-labiale superius plus</u> <u>Labiale superius-stomion</u>
14	Vermilion height index	<u>Labiale superius-stomion</u> <u>Stomion-labiale inferius</u>
15	Chin-mandible height index	<u>Sublabiale-gnathion</u> <u>Stomion-gnathion</u>
16	Upper face height-biocular width index	<u>Nasion-stomion</u> <u>Exocanthion (r)-exocanthion (l)</u>
17	Intercanthal-nasal width index	<u>Endocanthion (r)-endocanthion(l)</u> <u>Alare (r)-Alare (l)</u>
18	Nose-face height index	<u>Nasion-subnasale</u> <u>Nasion-gnathion</u>
19	Nose-mouth width index	<u>Alare (r)-alare (l)</u> <u>Chelion (r)-chelion (l)</u>
20	Upper lip-upper face height index	<u>Subnasale-stomion</u> <u>Nasion-stomion</u>
21	Upper lip-mandible height index	<u>Subnasale-stomion</u> <u>Stomion-gnathion</u>
22	Upper lip-nose height index	<u>Subnasale-stomion</u> <u>Nasion-subnasale</u>
23	Lower lip-face height index	<u>Stomion-sublabiale</u> <u>Subnasale-gnathion</u>
24	Lower lip-mandible height index	<u>Stomion-sublabiale</u> <u>Stomion-gnathion</u>
25	Lower lip-chin height index	<u>Stomion-sublabiale</u> <u>Sublabiale-gnathion</u>

Références

- [Adi97] Y. Adini, Y. Moses, S. Ullman, Face recognition: The problem of compensating for changes in illumination direction. *IEEE Trans. Patt. Anal. Mach. Intell.* 19, 721–732, 1997.
- [And97] M. C. De Andrade, G. Bertrand, A. Araujo, "Segmentation of microscopic images by flooding simulation : A catchment basins merging algorithm". *Proc. Of the ISI/SPIE Symp. On Electronic Imaging, Science & Technology, San-Jose-CA, USA, 1997.*
- [Ben03] S. Bengio. An Asynchronous Hidden Markov Model for Audio-Visual Speech Recognition. Dans S. BECKER, S. THRUN, et K. OBERMAYER, éditeurs, *Advances in Neural Information Processing Systems 15*, pages 1213–1220. MIT Press, 2003.
- [Ben05] S. Bennaceur, A.-S. Petavy-Blanc, J. Chauve, S. Louafi, J. Hanss. Morphologie céphalique humaine. *Données anthropométriques du vivant. EMCStomatologie*, 1, pages 85–103, 2005.
- [Ben06] B. BenAmor. Thèse Contributions à la modélisation et à la reconnaissance faciales 3D, Ecole Centrale de Lyon, 2006.
- [Bel97] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (7) 711–720, 1997.
- [Bel05] R. Belaroussi, L. Prevost, M. Milgram. Classifier combination for face localization in color images, M., *ICIAP'*, 2005.
- [Bes92] P.J. Besl, N.D. McKay. A Method for Registration of 3-D Shapes, *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2) :239–256, ISSN 0162-8828. 1992.
- [Beu79] S. Beucher, C. Lantuejoul, "Use of watersheds in contour detection". In *Proc. International Work-shop on Image Processing, Real-Time Edge and Motion Detection/Estimation, Rennes, sept.1979.*
- [Beu00] C. Beumier, M. Acheroy. Automatic 3D face authentication. In: *Proc. Image and Vision Computing, Vol. 18(4)*, pp. 315–321, 2000
- [Bio04] <http://www.biosecure.info/>
- [Bla01] D. Blackburn, M. Bone, P. J Phillips. "Face recognition vendor test 2000". Tech. rep. <http://www.frvt.org>, 2001.

- [Bla02] V. Blanz, R. Sami, V. Thomas, Face identification across different poses and illuminations with a 3d morphable model. Proc. IEEE International Conference on Automatic Face and Gesture Recognition, pp. 202–207, 2002.
- [Bla03] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Transactions on Pattern Analysis and Machine Intelligence 25, 1063–1074, 2003.
- [Bow06] K.W. Bowyer et al. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. Computer Vision and Image Understanding 101; 1–15; 2006.
- [Bra00] J. Brand, J. Mason, A comparative assessment of three approaches to pixel level human skin-detection, ICPR01, 1056–1059, 2000.
- [Bra05] G. Bradski, A. Kaehler and V. Pisarevsky, "Learning-Based Computer Vision with OpenCV." Intel Technology Journal, May 2005.
- [Bro01] D. Brown, I. Craw, J. Lewthwaite. A SOM based approach to skin detection with application in real time systems, BMVC01, 2001.
- [Bro03] 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.
- [Bru88]. V. Bruce. Recognizing faces. Lawrence Erlbaum Associates, London, U.K, 1988.
- [Bru93] R. Brunelli, T. Poggio. Face recognition: features versus templates. IEEE Trans. Pattern Anal. Mach. Intell, pp-1042–1062, 1993.
- [Car95] T. Carron, “Segmentation d’images couleur dans la base Teinte Luminance Saturation : approche numérique et symbolique”. Phd Thesis, Université de Savoie, Décembre 1995.
- [Cha97] C.C. Chang, L.L. Wang. A fast multilevel thresholding method based on lowpass and highpass filtering. Pattern Recognition Letters 18(14): 1469-1478, 1997.
- [Cha99] D. Chai, K.N. Ngan. Face segmentation using skin-color map in videophone applications, IEEE Trans. Circuits Syst. Video Technol.9 (4) 1999.
- [Cha03] 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.
- [Che04] S.C. Chen, J. Liu, Z.-H. Zhou, Making FLDA applicable to face recognition with one sample per person, Pattern Recognition 37 (7), 1553–1555, 2004.
- [Chu97] C.S. Chua, R. Jarvis. Point signature: A new representation for 3d object recognition. Internat. J. Computer Vision 25 (1), 63–85, 1997.
- [Chu00] C.S. Chua, F. Han, Y.K. Ho. 3D human face recognition using point signature. In: Proc. 4th IEEE Internat. Conf. on Automatic Face and Gesture Recognition (FG 2000) Grenoble, France, March, pp. 233–238, 2000.

- [Coc95] J.P. Cocquerez, S. Philipp, " Analyse d'image : filtrage et segmentation ", MASSON, Paris, France, 1995.
- [CNN02] CNN World News. Schiphol Backs Eye Scan Security. Available at <http://www.cnn.com/2002/WORLD/europe/03/27/schiphol.security/>, March 27, 2002.
- [Com60] J. Comas. Manual of Physical Anthropology. Charles C. Thomas, 1960.
- [Coo96] T.F. Cootes and C.J. Taylor. Locating Faces Using Statistical Feature Detectors. Proc. Second Int'l Conf. Automatic Face and Gesture Recognition, pp. 204-209, 1996.
- [Coo01] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6), 681–685, 2001.
- [Coo04] J.Cook, V.Chandran, S.Sridharan, C. Fookes. Face recognition from 3D data using iterative closest point algorithm and Gaussian mixture models. In: Proc. 2nd Internat. Symposium on 3D Data Processing, Visualization and Transmission (3DPVT 2004), Thessaloniki, Greece, 6–9 September, 2004.
- [Cos02] N.P. Costen, T.F. Cootes, C.J. Taylor. Compensating for ensemble specific effects when building facial models, Image Vision Comput. 20 673–682, 2002.
- [Cri04] D. Cristinacce, T.F. Cootes. A comparison of shape constrained facial feature detectors. Automatic Face and Gesture Recognition, Proceedings. Sixth IEEE, 375- 380, 2004.
- [Dau99] J. Daugman. Recognizing Persons by Their Iris Patterns. In A. K. Jain, R. Bolle, and S. Pankanti, editors, Biometrics: Personal Identification in a Networked Society, pp. 103-121, Kluwer Academic Publishers, 1999.
- [DeC98] D. DeCarlo, D. Metaxas, and M. Stone. An anthropometric face model using variational techniques. In SIGGRAPH, pages 67–74, 1998.
- [Dig78] H. Digabel, C.Lantuejoul. " Iterative algorithms". In Actes du Second Symposium Européen d'Analyse Quantitative des Microstructures en Sciences des Matériaux, Biologie et Medecine, Caen,4-7 October 1977 (1978).
- [Dem77] A. P. Dempster, N. M. Laird, D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm Journal of the Royal Statistical Society. Series B (Methodological), Vol. 39, No. 1, pp. 1-38,1977.
- [DeS95] L. C. De Silva, K. Aizawa, and M. Hatori, Detection and tracking of facial features by using a facial feature model and deformable circular template, IEICE Trans. Inform. Systems E78–D(9), 1195–1207, 1995.
- [Des04] Y. Desbois, R. Perrot, C. Debois. Incidence de l'occlusion dentaire lors d'une craniophotocomparaison : à propos d'un cas. Paleobios, Vol. 13, 2004.
- [DiZ86] S. Di Zenzo, "A note on the gradient of a multi-image", CVGIP : Graphics and image processing, pp 116-125, 1986.
- [Dud01] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification, second ed. John Wiley & Sons, 2001.

- [Duf05] S. Duffner, C. Garcia, "A Connexionist Approach for Robust and Precise Facial Feature Detection in Complex Scenes" , Fourth International Symposium on Image and Signal Processing and Analysis (ISPA 2005), Zagreb, Croatia, Septembre 2005.
- [Dui95] R.P.W. Duin. Small sample size generalization, in: G. Borgefors (Eds.), SCIA'95, Proceeding of the Ninth Scandinavian Conference on Image Analysis, vol. 2, Uppsala, Sweden, 6–9, pp. 957–964, 1995.
- [Edl06] R. Edler, P. Agarwal, D. Wertheim, D. Greenhill. The use of anthropometric proportion indices in the measurement of facial attractiveness. *The European Journal of Orthodontics* 28(3):274-281, 2006.
- [ECU] <http://www.some.ecu.au/~sphung>.
- [Ela01] A. Elad, R. Kimmel. "Bending Invariant Representations for. Surfaces," Proc. Computer Vision and Pattern Recognition (CVPR), 2001
- [Far80] L.G. Farkas, W. Bryson, J. Klotz. Is photogrammetry of the face reliable? *Plastic and Reconstructive Surgery* 66: 346 – 355, 1980.
- [Far87] L. G. Farkas. *Anthropometric Facial Proportions in Medicine*. Thomas Books, 1987.
- [Fau93] O. Faugeras, *Three-Dimensional Computer Vision: a Geometric Viewpoint*, MIT press, 1993.
- [Fer01] R. Féraud, O.J. Bernier, J.E. Viallet, M. Collobert. "A Fast and Accurate Face Detector Based on Neural Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 1, pp. 42-53, January, 2001.
- [FRGC02] "The Face Recognition Grand Challenge." <http://frvt.org/FRGC/>.
- [Fu81] S.K. Fu, J.K. Mu. A survey on image segmentation. *Pattern Recognit.* 13(1), 3-16, 1981.
- [Gar99] C. Garcia, G. Tziritas. Face detection using quantized skin color regions merging and wavelet packet analysis, *IEEE Trans. Multimedia* 1 (3) 264–277, 1999.
- [Gar04] C. Garcia, M. Delakis. Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection *IEEE Trans. PAMI*, vol. 26, no. 11, Nov, 2004.
- [Gok05] B. Gokberk, A.A. Salah, L. Akarun, Rank-based decision fusion for 3D shape-based face recognition, in: *International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA 2005)*, LNCS, vol. 3546, pp. 1019–1028, 2005.
- [Gom02] G. Gomez, E. Morales, Automatic feature construction and a simple rule induction algorithm for skin detection, *Proceedings of Workshop on Machine Learning in Computer Vision*, pp. 31–38, 2002.
- [Gre01] H. Greenspan, J. Goldberger, I. Eshet. Mixture model for facecolor modeling and segmentation, *Pattern Recognition Lett.* 22 (14) 1525–1536, 2001.

- [Gro02] R. Grossman, N. Kiryati, R. Kimmel. Computational surface flattening: a voxel-based approach, *IEEE Trans. PAMI*, 24, pp. 433-441, 2002.
- [Gro01]. R. Gross, J. Shi, J. Cohn. Quo Vadis Face Recognition?. Third Workshop on Empirical Evaluation Methods in Computer Vision, December, 2001.
- [Hal99] P. W. Hallinan, G. G. Gordon, A. L. Yuille, P. Giblin and D. Mumford. Two-and Three-dimensional Patterns of the Face, A K Peters, Ltd., Natick, MA, 1999.
- [He05] X. He, X. Yan, Y. Hu, p. Niyogi, H. Zhang. Face recognition using Laplacianfaces, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (3) 328–340, 2005.
- [Hei01] B. Heisele, T. Serre, M. Pontil, T. Poggio. Component-based face detection, in: *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 657–662, 2001.
- [Hes03] C. Heshner, A. Srivastava, G. Erlebacher. A novel technique for face recognition using range imaging. In *ISSPA*, pp. 201-204, 2003.
- [Hes04a] T. Heseltine, N. Pears, J. Austin. Three-dimensional face recognition: An eigensurface approach. In: *Proc. Internat. Conf. on Image Processing (ICIP 2004)*, October, Singapore, 2004a.
- [Hes04b] T. Heseltine, N. Pears, J. Austin. Three-dimensional face recognition: A fischersurface approach. In: *Proc. Image Analysis and Recognition: Internat. Conf. (ICIAR 2004)*, Porto, Portugal (September–October) 2004b.
- [Hor95] R. Horaud, O. Monga. *Vision par ordinateur*. Editions Hermès, 1995.
- [Hor97] T. Horprasert, Y. Yacoob, and L. S. Davis. An anthropometric shape model for estimating head orientation. In *3rd International Workshop on Visual Form*, Capri, Italy, May 1997.
- [Hrd39] A. Hrdlicka. *Practical Anthropometry*. Wister Institute of Anatomy and Biology, Philadelphia, 1939.
- [Hsu02] R.L. Hsu, M. Abdel-Mottaleb, A.K. Jain, Face detection in color images, *IEEE Trans. Pattern Anal. Machine Intell.* 24 (5) 696–706, 2002.
- [Hu89] G. Hu, G. Stockma. 3-D surface solution using structured light and constraint propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(4):390-402, 1989.
- [Hua03] J. Huang, B. Heisele, V. Blanz. Component-based face recognition with 3D morphable models, in: *Proceedings, International Conference on Audio- and Video-Based Person Authentication*, 2003.
- [Hus05] M.Husken, M.Brauckmann, S.Gehlen, and C. Vonder Malsburg. Strategies and benefits of fusion of 2D and 3D face recognition. In *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, (3)-174–174, 2005.
- [IBG] International Biometric Group, <<http://www.biometricgroup.com/>>.

- [Irf04] M.O. Irfanoglu, B. Gokberk, L. Akarun. 3D shape-based face recognition using automatically registered facial surfaces. In: Proc. 17th Internat. Conf. on Pattern Recognition (ICPR2004), Cambridge, pp. 183–186, 2004.
- [Isa56] E. J. Isaac , R. C. Singleton, Sorting by Address Calculation, Journal of the ACM (JACM), v.3 n.3, p.169-174, July 1956
- [IV2] IV2 website : <http://lsc.univ-evry.fr/techno/iv2/PageWeb-IV2.html>.
- [Jai82] A.K. Jain, B. Chandrasekaran. Dimensionality and sample size considerations in pattern recognition practice, in: P.R. Krishnaiah, L.N. Kanal (Eds.), Handbook of Statistics, vol. 2, pp. 835–855, 1982.
- [Jai87] A.K. Jain, B. Chandrasekaran. Dimensionality and sample size considerations in pattern recognition practice, in: P.R. Krishnaiah, L.N. Kanal (Eds.), Handbook of Statistics, vol. 2, North-Holland, Amsterdam, pp. 835–855, 1987.
- [Jai04] A.K. Jain, R. Arun, P. Salil. An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, Vol. 14, No. 1, January 2004.
- [Jen98] S.-H. Jeng, H. Y. M. Yao, C. C. Han, M. Y. Chern and Y. T. Liu. Facial Feature Detection Using Geometrical Face Model: An Efficient Approach. Pattern Recognition, 31(3), pp. 273-282, 1998.
- [Jon99] M. J. Jones, and J. M. Rehg. Statistical color models with application to skin detection. In *Proc. of the CVPR '99*, vol. 1, 274–280, 1999.
- [Jon02] M. J. Jones and J. M. Rehg. Statistical Color Models with Application to Skin Detection. *Int. J. Computer Vision* 46(1), pp. 81-96, 2002.
- [Jun04] H.C. Jung, B.W. Hwang, S.W. Lee. Authenticating corrupted face image based on noise model, Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004, 272pp.
- [Kak04] P. Kakumanu, S. Makrogiannis, R. Bryll, S. Panchanathan, and N. Bourbakis. Image chromatic adaptation using ANNs for skin color adaptation, Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence, ICTAI04.
- [Kak06] P. Kakumanu. A face detection and facial expression recognition method applicable to assistive technologies and biometrics, PhD Dissertaion, CSE Department, Wright State University, 2006.
- [Kan73] T. Kanade. “Picture Processing by Computer Complex and Recognition of Human Faces,” PhD thesis, Kyoto Univ., 1973.
- [Kel70] M.D. Kelly. Visual identification of people by computer, Technical Report AI-130, Stanford AI Project, Stanford, CA, 1970.
- [Kep02] B. Kepenekci, F.B. Tek, G. Bozdagi Akar. Occluded face recognition based on Gabor wavelets, ICIIP 2002, Rochester, NY, MP-P3.10. , September 2002.
- [Kim03] T.K. Kim, H. Kim, W. Hwang, S.C. Kee, J. Kittler. Independent component analysis in a facial local residue space. In: Proc. IEEE Computer Society

- Conference on Computer Vision and Pattern Recognition, Vol. 1 (June), pp. 579–586, 2003.
- [Kit98] J. Kittler, M. Hatef, R. Duin, and J. Matas, On Combining Classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, 226-239, Mar. 1998.
- [Koh97] T. Kohonen. *Self-Organizing Map*, second ed., Springer, Berlin, 1997.
- [Kot97] C. Kotropoulos and I. Pitas. Rule-Based Face Detection in Frontal Views. *Proc. Int'l Conf. Acoustics, Speech and Signal Processing*, vol. 4, pp. 2537-2540, 1997.
- [Lad93] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, Malsburg von der, R. Wurtz. Distortion invariant object recognition in the dynamic link architecture, *IEEE Trans. Comput.* 42 (3) (1993) 300–311.
- [Lan95] A. Lanitis, C.J. Taylor, T.F. Cootes. Automatic face identification system using flexible appearance models, *Image Vis. Comput.* 13 (5) 393–401, 1995.
- [Law97] S. Lawrence, C.L. Giles, A. Tsoi, A. Back. Face recognition: a convolutional neural-network approach, *IEEE Trans. Neural Networks* 8 (1) 98–113, 1997.
- [Le04] H.S. Le, H. Li. Recognizing frontal face images using hidden Markov models with one training image per person, *Proceedings of the 17th International Conference on Pattern Recognition (ICPR04)*, vol. 1, pp. 318–321, 2004.
- [Lee96] T.S. Lee, Image representation using 2-d Gabor wavelets, *IEEE Trans. Pattern Anal. Mach. Intell.* 18 (10) 959–971, 1996.
- [Lee02] J.Y. Lee, S. I. Yoo. An elliptical boundary model for skin color detection. In *Proc. of the International Conference on Imaging Science, Systems, and Technology*, 2002.
- [Lee05] Y. Lee, H. Song, U. Yang, H. Shin, and K. Sohn. Local feature based 3d face recognition. In *Audio- and Video-based Biometric Person Authentication*, 2005 International Conference on, LNCS, 3546-909–918, 2005.
- [Li99] S.Z. Li, J. Lu. Face recognition using the nearest feature line method, *IEEE Trans. Neural Networks* 10 (2) 439–443, 1999.
- [Li04] S.Z. Li, Z.Q. Zhang. FloatBoost Learning and Statistical Face Detection, *PAMI(26)*, No. 9, pp. 1112-1123, 2004.
- [Li05] S.Z. Li, A.K. Jain. *Handbook Of Face Recognition*. Springer, 2005.
- [Lin99] C. Lin and J. Wu. Automatic facial feature extraction by genetic algorithm. *IEEE Transactions on Image Processing*, 8(6):834--845, June 1999.
- [Liu00] C. Liu, H. Wechsler, Evolutionary pursuit and its application to face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (6) 570–582, 2000.

- [Lu03] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, Face recognition using kernel direct discriminant analysis algorithms, *IEEE Trans. Neural Networks* 14 (1) 117–126, 2003.
- [Man92] B.S. Manjunath, R. Chellappa, C.V.D. Malsburg, A feature based approach to face recognition, in: *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 373–378, 1992.
- [Mar01] A. Martinez, A.C. Kak. PCA versus LDA. *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (2) 228–233, 2001.
- [Mar02] A.M. Martinez, Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (6) (2002) 748–763.
- [Mar03] B. Martinkauppi, M. Soriano, M. Pietikäinen, Detection of skin color under changing illumination: a comparative study, 12th International Conference on Image Analysis and Processing, 2003.
- [Mar98] A.M. Martínez, R. Banavente. The AR face database. Tech. Report 24 CVC Barcelone, Espagne, June 1998.
- [Mes99] K. Messer, J. Matas, J. Kittler, J. Luetttin, and G. Maitre. XM2VTSDB: The Extended M2VTS Database. In *Proceedings, International Conference on Audio- and Video-Based Person Authentication*. 72–77, 1999.
- [Mog97] B. Moghaddam, A. Pentland, Probabilistic visual learning for object representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (7) 696–710, 1997
- [Mor03] A.B. Moreno, A. Sanchez, J.F. Velez Diaz. Face recognition using 3D surface-extracted descriptors. In: *Proc. Irish Machine Vision and Image, (IMVIP'03)*, September, 2003.
- [Nik98] A. Nikolaidis, and I. Pitas, "Facial feature extraction and determination of pose," in *Proc. of the 1998 NOBLESSE Workshop on Nonlinear Model Based Image Analysis*, 1998.
- [Oja02] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 971–987, 2002.
- [Oli61] G. Olivier. *Morphologie et types humains*. Vigot Frères, 1961, 171 pages.
- [OTo93] A.J. O'Toole, H. Abdi, Low-dimensional representation of faces in higher dimensions of the face space, *Opt. Soc. Am.* 10 (3), 405–411, 1993.
- [Ots79] N.A. Otsu. thresholding selection method from gray-level histogram. *IEEE Trans. Syst. Man Cybernet.* 9(1), 62-66, 1979.
- [Pee03] P. Peer, J. Kovac, F. Solina. Human skin colour clustering for face detection. In submitted to EUROCON 2003 – International Conference on Computer as a Tool, 2003.
- [Pen94] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *Proceedings, IEEE Conference*, 1994

- [Pen96] P. Penev, J. Atick, Local feature analysis: a general statistical theory for object representation, *Netw.: Comput. Neural Syst.* 7 477–500, 1996.
- [Per96] R. Perrot. “Use of Anthropological Methods in the Identification of unknown Individuals : Human Remains and Armed Robbers”. 14th Meeting of the International Association of Forensic Sciences , Tokyo, Japan, 1996.
- [Phi98a] P.J. Phillips, Support vector machines applied to face recognition, *Adv. Neural Inform. Process. Syst.* 11 (03) 809, 1998.
- [Phi98b] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image Vis. Comput.* 16, 295–306, 1998b.
- [Phi00] P.J. Phillips, H. Hyeonjoon, S. Rizvi, P. Rauss. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vo. 22, No. 10, Octobre 2000.
- [Phi03] 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.
- [Phi07] P.J. Phillips, and al. FRVT2006 and ICE2006 Large Scale Results. NIST Report, March 2007
- [Por00] G. Porter, G. Doran. An anatomical and photographic technique for forensic facial identification. *Forensic Science International* 114. Pages: 97-105, 2000.
- [Pra07] S. Prabhakar, J. Kittler, D. Maltoni, L. O’Gorman, and T. Tan. Introduction to the Special Issue on Biometrics: Progress and Directions , *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, April 2007.
- [Pri05] J.R. Price, R. Jeffery, T.F. Gee, F. Timothy. Face recognition using direct, weighted linear discriminant analysis and modular subspaces, *PR(38)*, No. 2, pp. 209-219, February 2005.
- [Rau91] S.J. Raudys, A.K. Jain, Small sample size effects in statistical pattern recognition: recommendations for practitioners, *IEEE Trans. Pattern Anal. Mach. Intell.* 13 (3) 252–264, 1991.
- [Riz98] S. A. Rizvi, P. J. Phillips, H.A. Moon. Verification protocol and statistical performance analysis for face recognition algorithms. In *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*. 833–838, 1998.
- [Roe 98] G. Roethenbaugh. “An Introduction to Biometrics and General History”, *Biometrics Explained*, Section 1, 1998.
- [Row98] H. A. Rowley, S. Baluja, et T. Kanade, "Neural Network based Face Detection", *IEEE Trans, Pattern Anal. Mach, Intell*, 23-38, January 1998
- [Sam93] F. Samaria, Face segmentation for identification using hidden Markov models, in: *British Machine Vision Conference*, BMVA Press, pp. 399–408, 1993.
- [Sam94] F. Samaria, A. Harter, Parameterisation of a stochastic model for human face identification, in: F.L. Sarasota (Ed.), *Proceedings of Second IEEE*

- Workshop on Applications of Computer Vision, pp. 138–142, December 1994.
- [Sch00] H. Schneiderman and T. Kanade. “A Statistical Method for 3D Object Detection Applied to Faces and Cars”. PhD thesis, RI, 2000.
- [Sch00a] H. Schneiderman and T. Kanade, “A Statistical Method for 3D Object Detection Applied to Faces and Cars” Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 746-751, 2000.
- [She81] J.W. Shepherd, G.M. Davidies, H.D. Ellis, Studies of cue saliency, in: G.M. Davies, H.D. Ellis, J.W. Shepherd (Eds.), *Perceiving and Remembering aces*, Academic Press, London, UK, 1981.
- [Sim02] T. Sim, S. Baker, M. Bsat, The CMU pose illumination and expression (PIE) database, in: *Proceedings of the Fifth International Conference on Automatic Face and Gesture Recognition*, 2002.
- [Sin94] P. Sinha, “Object Recognition via Image Invariants: A Case Study,” *Investigative Ophthalmology and Visual Science*, vol. 35, no. 4, pp. 1735-1740, 1994.
- [Sin95] P. Sinha, “Processing and Recognizing 3D Forms,” PhD thesis, Massachusetts Inst. of Technology, 1995.
- [Sin06] P. Sinha, B. Balas, Y. Ostrovsky, R. Russell. *Face Recognition by Humans: 19 Results all Computer Vision Researchers Should Know About*. *Proceedings of the IEEE, 1948-1962*, 2006.
- [Sir87] L. Sirovich, M. Kirby, Low-dimensional procedure for the characterization of human faces, *J. Opt. Soc. Am. A* 4 (3) (1987) 519–524.
- [Sne05] R. Snelick, U. Uludag, A. Mink, M. Indovina, A. Jain. Large-Scale Evaluation of Multimodal Biometric Authentication Using State-of-the-Art Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 450-455, Mar., 2005.
- [Sob96] K. Sobottka, I. Pitas. Extraction of facial regions and features using color and shape information. In: *Proc. 13th Internat. Conf. Pattern Recognition*, Vienna, Austria, pp. 421-425, 1996.
- [Sob98] K. Sobottka, I. Pitas, A novel method for automatic face segmentation, facial feature extraction and tracking, *Signal Process. Image Commun.* 12, 263–281, 1998.
- [Soh06] A.S.M. Sohail, P. Bhattacharya. Localization of Facial Feature Regions Using Anthropometric Face Model. In: *First International Conference on Multidisciplinary Information Sciences and Technologies*, 2006.
- [Tan06] X. Tana, C. Songcan. Face recognition from a single image per person: A survey. *Pattern Recognition*, 2006.
- [Tan05] X. Tan, S.C. Chen, Z.-H. Zhou, F. Zhang, Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft kNN ensemble, *IEEE Trans. Neural Networks* 16 (4) 875–886, 2005.

- [Ter98a] J.C. Terrillon, M. David, and S. Akamatsu, "Automatic Detection of Human Faces in Natural Scene Images by Use of a Skin Color Model and Invariant Moments," Proc. Third Int'l Conf. Automatic Face and Gesture Recognition, pp. 112-117, 1998.
- [Ter98b] J.C. Terrillon, M. David, and S. Akamatsu. "Detection of Human Faces in Complex Scene Images by Use of a Skin Color Model and Invariant Fourier-Mellin Moments," Proc. Int'l Conf. Pattern Recognition, pp. 1350-1355, 1998.
- [Tur91] M. Turk and A. Pentland. "Eigenfaces for Recognition," J. Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.
- [Tsa87] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lens" IEEE J. Robotics Automation, vol. RA-3, no. 4, pp. 323-344, 1987.
- [Tsa95] D.M. Tsai. A fast thresholding selection procedure for multimodal and unimodal histograms. Pattern Recognit. Lett. 16(6), 653-666,1995.
- [Tsa04] F. Tsalakanidou, S. Malassiotis, M. Strintzis, Integration of 2D and 3D images for enhanced face authentication, in: Sixth International Conference on Automated Face and Gesture Recognition, pp. 266–271, 2004.
- [Ver04] F. Ventura, A. Zacheo, A. Ventura. Computerised anthropomorphometric analysis of images : case report. Forensic Science International 146s, pages S211-S213, 2004.
- [Vet97] T. Vetter and T. Poggio, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 733–742, 1997.
- [Vez03] V. Vezhnevets, V. Sazonov, A. Andreeva. A Survey on Pixel-Based Skin Color Decton Techniques, Proc. Graphicon-2003 pp. 85-92, Moscow, Russia, September 2003.
- [Vin91] L. Vincent, P. Soille, "Watersheds in digital space : an efficient algorithm based on immersion simulation". IEEE Trans on PAMI, 13 (6), pp 563-598, 1991.
- [Vio01] P. Viola, M. Jones. Rapid object detection using a boosted cascade of simple features. In Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, 2001.
- [Vis05] M. Visani, C. Garcia, J.M. Jolion, "Bilinear Discriminant Analysis for Face Recognition". Dans Proceedings of the International Conference on Advances in Pattern Recognition (ICAPR 2005), volume 2, pages 247-256, Août 2005.
- [Wan02] Y. Wang, C. Chua, and Y. Ho. Facial feature detection and face recognition from 2D and 3D images. Pattern Recognition Letters, 23:1191–1202, 2002.
- [Wan03] X. Wang, X. Tang, Unified subspace analysis for face recognition, Proceedings of the Ninth IEEE International Conference on Computer Vision, 2003, pp. 679–686.

- [Wan05] J. Wang, K.N. Plataniotis, A.N. Venetsanopoulos, Selecting discriminant eigenfaces for face recognition, *Pattern Recognition Lett.* 26 (10) 1470–1482, 2005.
- [War89] E. L.Wartz, A. Shaw, E. Wolfson. A numerical solution to the generalized mapmaker's problem: flattening nonconvex polyhedral surfaces, *IEEE Trans. PAMI*, 11, pp. 1005-1008, 1989.
- [Wis97] L. Wiskott, R. Fellous, N. Kruger, C. von Malsburg. Face recognition by elastic bunch graph matching, *IEEE Trans. Pattern Anal. Mach. Intell.* 775–779, July 1997.
- [Wu02] J. Wu, Z.-H. Zhou. Face recognition with one training image per person, *Pattern Recognition Lett.* 23 (14) (2002) 1711–1719.
- [Yan94] G. Yang and T. S. Huang. “Human Face Detection in Complex Background,” *Pattern Recognition*, vol. 27, no. 1, pp. 53-63, 1994.
- [Yan98a] W. Yang, A. Waibel Lu. Skin-color modeling and adaptation, *ACCV98*, 1998.
- [Yan98b] M.H. Yang, N. Ahuja. Detecting human faces in color images, *ICIP98*, 1998.
- [Yan99] M.H. Yang, N. Ahuja. Gaussian Mixture model for human skin color and its application in image and video databases, *Proceedings of SPIE: Conference on Storage and Retrieval for Image and Video Databases*, vol. 3656, pp. 458–466, 1999.
- [Yan02] Ming-Hsuan Yang, David J. Kriegman et Narendra Ahuja. Detecting faces in images : A survey. Dans *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 24(1), pages 34–58, 2002.
- [Yan04] J.Yang, D. Zhang, A.F. Frangi, J.Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 131–137,2004.
- [Yow97] K.C. Yow and R. Cipolla, “Feature-Based Human Face Detection,” *Image and Vision Computing*, vol. 15, no. 9, pp. 713-735, 1997.
- [Yui92] A. Yuille, P. Hallinan, and D. Cohen, “Feature Extraction from Faces Using Deformable Templates,” *Int’l J. Computer Vision*, vol. 8, no. 2, pp. 99-111, 1992.
- [Xu06] Z. Xu, M. Zhu. ”Color-based skin detection: survey and evaluation”. *Multi-Media Modelling Conference Proceedings*, 2006 12th International. Hangzhou, China; 2006.
- [Zha97] J. Zhang, Y. Yan, M. Lades, Face recognition: eigenface, elastic matching and neural nets, *Proc. IEEE* 85 (9) (1997) 1423–1435.
- [Zha99] W. Zhao, R. Chellappa, P.J. Phillips, Subspace linear discriminant analysis for face recognition, *Technical Report CAR-TR-914*, Center for Automation Research, University of Maryland, 1999.
- [Zig02] G. Zigelman, R. Kimmel, N. Kiryati. Texture mapping using surface flattening via multi-dimensional scaling, *IEEE Trans. Visualization and Comp. Graphics*, 8, pp. 198-207, 2002.