

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université 08 Mai 1945 - Guelma

Faculté des Sciences et de la Technologie

Département Electronique et Télécommunications

LABORATOIRE PROBLEMES INVERSES, MODELISATION,
INFORMATION ET SYSTEMES (PIMIS)



THÈSE

Présentée en vue de l'obtention du diplôme de:

Doctorat 3^{ème} Cycle en Electronique

Option : Images et Signaux Biométriques

Intitulée :

**Identification Biométrique par Descripteurs de Texture
Locaux: Application au Visage & Oreille**

Présentée par : **Amir BENZAOUI**

Devant le jury:

Abdelkrim MOUSSAOUI	Prof des Universités	Univ. Guelma	Président
Abdelhani BOUKROUCHE	Prof des Universités	Univ. Guelma	Rapporteur
Amina SERIR	Prof des Universités	USTHB Alger	Examinatrice
Mohamed Tarek KHADIR	Prof des Universités	Univ. Annaba	Examinateur
Hamid SERIDI	Prof des Universités	Univ. Guelma	Examinateur
Imed BOUCHRIKA	Maître de conférences. A	Univ. Souk Ahras	Examinateur

Année Universitaire 2014 / 2015

Soutenue le: 18/05/2015

Dédicaces

Je dédie ce travail :

À ceux ; qui m'ont tant donné sans rien demander, qui m'ont toujours offert leur soutien m'ont épargné aucun effort pour m'aider, qui m'ont appris mes principes, à qui tous les mots ne suffisent pas pour les remercier :

« Mes très chers parents »

À mes frères avec mes souhaits de bonheur et de prospérité.

À ma fiancée pour sa compréhension et son soutien réel.

À mon encadreur pour son aide précieuse

À mes collègues

Aux membres du laboratoire PI:MIS

À toute ma famille

À tous mes amis

À toute personne, qui m'a aidée à réaliser ce modeste travail, de proche ou de loin.

Remerciements

Je remercie en premier lieu mon grand DIEU qui m'a donné à la fois le courage, la volonté, et la patience afin d'élaborer cette thèse de recherche scientifique.

- *Je tiens à remercier, mon directeur de thèse Monsieur **Abdelhani BOUKROUCHE**, Professeur à l'université 08 Mai 1945 à Guelma et directeur de notre laboratoire **PI:MIS**, pour l'encadrement de mon travail et pour son encouragement, ainsi que son soutien tout au long de la thèse. Je le remercie pour tout son aide. Son enthousiasme et sa patience ont beaucoup facilité et agrémenté mon travail. Il a été toujours disponible pour répondre aux questions que je lui posais. Ses remarques m'ont permis de faire progresser ce travail.*
- *Je tiens également à remercier les membres du jury qui m'ont fait l'honneur de bien vouloir évaluer mon travail, et plus précisément :*
 - ✓ *Monsieur **Abdelkrim MOUSSAOUI**, Professeur à l'Université de Guelma, pour l'honneur qu'il m'a fait, en acceptant la présidence de ce jury.*
 - ✓ *Madame **Amina SERIR**, Professeur à l'USTHB d'Alger, Monsieur **Mohamed Tarek KHADIR**, Professeur à l'Université d'Annaba, Monsieur **Hamid SERIDI**, Professeur à l'Université de Guelma, et Monsieur **Imed BOUCHRIKA**, Maître de conférences à l'Université de Souk Ahras, d'avoir accepté de juger le présent document.*
- *Un remerciement spécial est aussi adressé aux membres du **Center for Machine Vision (CMV)** de l'université d'Oulu (Finland), en particulier Professeur **Abdenour HADID** pour avoir ponctué mon travail de discussions fructueuses. Je le remercie très sincèrement pour son extrême gentillesse, son aide et ses encouragements durant toute la période de préparation de cette thèse.*

- *Je tiens aussi à exprimer mes très sincères remerciements à tous les collègues du laboratoire **PI:MIS** pour leur soutien. Sans oublier de remercier Monsieur **Hocine BOUROUBA**, Maître de conférences à l'université de Guelma, qui a accepté de donner des heures de leur temps libre pour me permettre de réaliser plusieurs expérimentations de cette thèse.*
- *Je remercie également tous les autres collègues de notre laboratoire avec lesquelles j'ai pu avoir de nombreux échanges, et évoluer mon esprit de débat scientifique, en leurs espérant une bonne continuation et une bonne chance.*
- *J'adresse un grand MERCI à toute l'équipe du **Système National de Documentation en Ligne (SNDL)** qui a mis à la disposition de tous les chercheurs Algériens une immense base de documentation Multidisciplinaires gratuite (ouvrages, thèses, rapports, communications, revues et bases de données scientifiques, etc.).*

Résumé

Identification Biométrique par Descripteurs de Texture Locaux: Application au Visage & Oreille

Les systèmes biométriques offrent une identification automatique de l'identité, basés sur des mesures biologiques de l'individu. Nous nous sommes intéressés dans cette étude à deux modalités biométriques, qui sont le *visage* et l'*oreille*. Tout d'abord, la reconnaissance par visage possède plusieurs avantages par rapport aux autres technologies; elle est naturelle, non-intrusive et facile à utiliser. Cependant, une identification rapide et précise par visage présente toujours un certain nombre de défis et de difficultés, particulièrement dans les environnements non-contrôlés. Par ailleurs, l'oreille humaine est considérée comme une modalité très récente dans ce domaine, elle peut être utilisée comme une pièce complémentaire au visage. L'identification par oreille reste jusqu'à présent en cours d'exploration, son avantage principal réside dans sa richesse d'information et sa stabilité au cours du temps.

Dans le cadre de cette thèse, nous proposons une approche d'extraction des caractéristiques basée principalement sur les descripteurs de texture locaux, dans le but d'identifier les individus à partir des images d'intensités 2D, de visages ou d'oreilles, dans des conditions non-contrôlées; en effet, quatre descripteurs ont été testés et comparés, à savoir: LBP (*Local Binary Pattern*), 1DLBP (*One Dimensional Local Binary Pattern*), LPQ (*Local Phase Quantization*) et BSIF (*Binarized Statistical Image Features*). Cette approche a la capacité à dériver facilement des modèles de caractéristiques efficaces qui combinent dans un vecteur de données unique la forme globale de l'objet analysé et la texture locale de son apparence; elle imite, d'une façon similaire, les capacités de l'être humain en reconnaissance d'objets, par sa particularité en codification des composantes à basses et hautes fréquences qui contribuent à la description des caractéristiques locales et globales respectivement.

Mots clés : Biométrie, visage, oreille, identification, descripteurs de texture, LBP, 1DLBP, LPQ, BSIF

Table des Matières

Dédicaces	ii
Remerciements.....	iii
Résumé.....	v
Table des Matières.....	vi
Liste des tableaux	xi
Liste des figures	xiii
Introduction Générale.....	17
Chapitre 1 Introduction à la Biométrie et aux Systèmes Biométriques.....	22
1.1 Introduction	22
1.2 Généralités et notions de bases en biométrie.....	22
1.3 Architecture fonctionnelle d'un système biométrique.....	25
1.3.1 Fonctionnement interne d'un système biométrique.....	26
1.3.2 Vérification et Identification.....	28
1.4 Performances d'un système biométrique.....	30
1.4.1 Test de vérification	30
1.4.2 Test d'identification.....	34
1.5 Applications de la biométrie.....	35
1.6 Caractéristiques de la biométrie.....	36
1.7 Considération.....	43
1.8 Biométrie cachée	45
1.8.1 Utilisation des images MRI en biométrie	46
1.8.2 Utilisation des images X-ray en biométrie	48

1.9	Conclusion	50
Chapitre 2	Reconnaissance Biométrique par Visages	51
2.1	Introduction	51
2.2	Le visage: le moyen naturel pour la reconnaissance humaine.....	51
2.3	Inspirations neuropsychologiques en reconnaissance de visages.....	53
2.4	Reconnaissance de visages dans le domaine de l'ingénierie.....	59
2.4.1	Etapes clés pour la reconnaissance: Détection de visages et extraction des caractéristiques.....	61
2.4.2	Segmentation / Détection de visages (Sommaire)	61
2.4.3	Extraction des caractéristiques (Sommaire et méthodes)	63
2.4.4	Traitement de l'illumination	72
2.4.5	Représentation numérique des visages humains.....	74
2.4.6	Reconnaissance de visages à distance	75
2.5	Défis ouverts et futures applications.....	76
2.5.1	Reconnaissance de visages et les occlusions	76
2.5.2	Reconnaissance de visages hétérogènes	78
2.5.3	Reconnaissance de visages et le vieillissement	79
2.6	Conclusion	80
Chapitre 3	Reconnaissance Biométrique par Oreilles.....	82
3.1	Introduction	82
3.2	Caractéristiques de l'oreille : Approche biologique	82
3.2.1	Anatomie de l'oreille humaine	83
3.2.2	Embryologie de l'oreille	84
3.2.3	Anomalies extérieures de l'oreille et les composantes sensibles aux variations	86
3.3	Système générique de reconnaissance automatique par oreilles	88

3.3.1	Détection de l'oreille	89
3.3.2	Détection de l'oreille	90
3.3.3	Extraction des caractéristiques / Réduction des paramètres	90
3.4	Approches de reconnaissance biométrique par oreilles.....	91
3.4.1	Approches basées sur les transformations	93
3.4.2	Approches locales	97
3.4.3	Approches hybrides	100
3.4.4	Approches statistiques / holistiques.....	102
3.5	Défis ouverts et futures application	106
3.5.1	Localisation automatique de l'oreille	106
3.5.2	Occlusion et les variations de la pose	107
3.5.3	Évolutivité des systèmes de reconnaissance par oreilles	108
3.5.4	Comprendre la symétrie et le vieillissement.....	108
3.6	Conclusion.....	109
Chapitre 4 Caractérisation Biométrique par Descripteurs de Texture Locaux		110
4.1	Introduction	110
4.2	Motivation	110
4.3	Notions fondamentales sur l'analyse de texture	111
4.3.1	Définition de la texture	111
4.3.2	Catégorisation des descripteurs de texture.....	112
4.3.3	Problèmes d'analyse de texture	113
4.3.4	Description de texture	116
4.3.5	Descripteurs de texture locaux.....	117
4.4	Motif binaire local (LBP: Local Binary Pattern).....	118
4.4.1	Dérivation de l'opérateur LBP génétique	121

4.4.2	Utilisation du LBP en cratérisation biométrique	123
4.5	LBP unidimensionnel (1DLBP)	125
4.6	Quantification par phase locale (LPQ: Local Phase Quantization)	127
4.7	Caractéristiques statistiques et binarisées de l'image (BSIF : Binarized Statistical Image Features)	130
4.8	Conclusion	133
Chapitre 5 Implémentation et Etudes Expérimentales		134
5.1	Introduction	134
5.2	Représentation et reconnaissance biométrique à base de descripteur de texture locaux	134
5.3	Etude expérimentale	136
5.3.1	Bases d'images	136
5.3.2	Expérimentation#1 (Effet du classificateur et le niveau de la décomposition)	141
5.3.3	Expérimentation#2 (Effet des paramètres de l'opérateur BSIF)	147
5.3.4	Expérimentation #3 (Effet de l'information rigide / semi-rigide)	150
5.3.5	Comparaison	154
5.4	Conclusion	156
Conclusion et Perspectives.....		157
Liste des Publications, Communications Indexées, et Evènements Scientifiques		161
Références Bibliographiques		162
Annexe A Notions fondamentales sur la classification, les k-plus proches voisins (k-NN) et les séparateurs à vaste marge (SVM)		185
	FERET	192
	The Facial Recognition Technology Database	192
	FRGC.....	192
	Face Recognition Grand Challenge Database	192

IIT Delhi	192
Indian Institute of Technology Delhi	192
IITK	192
Indian Institute of Technology Kanpur	192

Liste des tableaux

Tableau 1. 1 : Comparaison entre les modalités biométriques.....	42
Tableau 5. 1 : Distribution des images entre l'apprentissage et le test en utilisant deux images par personne en apprentissage.....	140
Tableau 5. 2 : Distribution des images entre l'apprentissage et le test en utilisant une seule image par personne en apprentissage.	141
Tableau 5. 3 : Taux d'identifications au rang-1 avec différents classificateurs et niveau de décomposition appliqués sur la base d'images ORL.....	142
Tableau 5. 4 : Taux d'identifications au rang-1 avec différents classificateurs et niveau de décomposition appliqués sur la base d'images AR.	142
Tableau 5. 5 : Taux d'identifications au rang-1 en utilisant deux images par personne en apprentissage avec différents classificateurs et niveau de décomposition appliqués sur la base d'images IIT Delhi-1.....	143
Tableau 5. 6 : Taux d'identifications au rang-1 en utilisant deux images par personne en apprentissage avec différents classificateurs et niveau de décomposition appliqués sur la base d'images IIT Delhi-2.....	143
Tableau 5. 7 : Taux d'identifications au rang-1 en utilisant deux images par personne en apprentissage avec différents classificateurs et niveau de décomposition appliqués sur la base d'images USTB.	144
Tableau 5.8 : Taux d'identifications au rang-1 en utilisant une seule image par personne en apprentissage avec différents classificateurs et niveau de décomposition appliqués sur la base d'images IIT Delhi-1.....	144
Tableau 5. 9 : Taux d'identifications au rang-1 en utilisant une seule image par personne en apprentissage avec différents classificateurs et niveau de décomposition appliqués sur la base d'images IIT Delhi-2.....	145

Tableau 5. 10 : Taux d'identifications au rang-1 en utilisant une seule image par personne en apprentissage avec différents classificateurs et niveau de décomposition appliqués sur la base d'images USTB.	145
Tableau 5. 11 : Taux d'identifications au rang-1 en utilisant tous les paramètres existants de l'opérateur BSIF appliqués sur la base d'images IIT Delhi-2, niveau-3 de la projection horizontale-verticale, 1-NN classificateur et deux images par personne en apprentissage.	148
Tableau 5. 12 : Taux d'identifications au rang-1 en utilisant les paramètres optimaux de l'opérateur BSIF (filtre de taille 15×15 pixels et 12 bits).	148
Tableau 5. 13 :Taux d'identifications au rang-1 en utilisant les informations rigides et semi-rigides appliqués sur les différentes bases d'images.	153
Tableau 5. 14 : Taux d'identifications au rang-1 en utilisant le " <i>Protocole B</i> " appliqué sur la base d'images AR.	153
Tableau 5. 15 : Sommaire des travaux connexes et récents en identification des individus par visages.	154
Tableau 5. 16 : Sommaire des travaux connexes et récents en identification des individus par oreilles.	155

Liste des figures

Figure 1. 1 : Exemples de modalités biométriques (physiologiques et comportementales)..25	25
Figure 1.2 : Architecture fonctionnelle d'un système biométrique. Diagramme des processus d'apprentissage, vérification et identification.....29	29
Figure 1. 3 : Distribution des scores de similarités des imposteurs et authentiques en mode vérification. Illustration de détermination du point d'équivalence des erreurs dans un système biométrique.....33	33
Figure 1.4 : Courbe ROC « Receiver Operating Characteristic ».33	33
Figure 1.5 : Courbe CMC.35	35
Figure 1.6 : Deux images MRI du cerveau humain qui montrent une différence visuelle entre deux individus.46	46
Figure 1.7 : Biométrie cachée par images 2D du cerveau humain..47	47
Figure 1.8 : Biométrie cachée en 3D..48	48
Figure1. 9 : Biométrie cachée appliquée sur les images X-ray.49	49
Figure1.10 : Biométrie cachée par les images X-ray appliqué au corps humain.49	49
Figure 2. 1 : Les régions d'activation du cerveau répondants principalement au stimulus de visage (motifs rouge au jaune) ou à des photos de maisons (motifs bleu).55	55
Figure 2.2 : L'illusion de <i>Thatcher</i> montre que les distorsions flagrantes sur les informations de base du visage sont rarement remarquées lorsque le visage est inversé.56	56
Figure 2. 3 : Exemple d'une application de type " <i>Mugshot-ID</i> ".57	57
Figure 2.4 : Configuration générique d'un système de reconnaissance automatique de visages.60	60
Figure 2. 5 : Composantes principales définies par la PCA (La décomposition de <i>Karunen Lowe</i>) et par l'analyse en composante indépendante (ICA).....67	67
Figure 2. 6 : Extraction des points de repère par l'algorithme EBGM.....69	69

Figure 2.7 : Quelques modalités d'imageries qui démontrent des hétérogénéités alternatives.	79
Figure 2.8 : Echantillons de visages de deux personnes dans le vieillissement.	80
Figure 3.1 : Anatomie de l'oreille humaine.....	83
Figure 3.2 : Présentation des variations interindividuelles entre les composantes de l'oreille.	84
Figure 3.3 : (a) Image d'un micrographe électronique d'un embryon humain âgé de six semaines. (b) Six hillocks auriculaires sont bien illustrés dans ce schéma (c) Voûtes mandibulaires et hyoïdes.....	85
Figure 3.4 : Différentes positions finales des hillocks, suggérées par différents auteurs.	85
Figure 3.5 : Différentes positions finales des hillocks, suggérées par différents auteurs.	88
Figure 3.6 : Exemple de normalisation de l'oreille dans la même direction et taille.	90
Figure 3.7 : Système d' <i>Iannarelli</i> en identification manuelle par oreilles.	92
Figure 3.8 :Étapes utilisées pour créer le modèle d'oreille basé sur le graphe d'adjacences.	92
Figure 3.10 : Exemple d'extraction des lignes du champ de force à partir d'une image d'oreille.	94
Figure 3.11 : (a) Repères SIFT d'une image d'oreille (b) Crus d'hélicé détecté par SIFT dans quatre images différentes.....	98
Figure 3.12 : Représentation symbolique de la méthode des cercles concentriques.....	99
Figure 3.13 : Caractéristiques géométriques sélectionnées par la méthode de <i>Sibai et al.</i> (2013) [206].....	100
Figure 3.14 : Exemple d'identification par la méthode de <i>Jeges et Mate</i> (2007) [209].	101
Figure 4.1 : Exemples de textures: (a) base d'images <i>Brodatz</i> (b) base d'images <i>KTH-TIPS2</i>	112
Figure 4.2 : Exemple de classification de texture.	114

Figure 4. 3 : Exemple de segmentation de texture.	115
Figure 4. 4 : Exemple d'extraction des caractéristiques en utilisant l'histogramme de l'opérateur LBP basique: Application sur une image entière de visage.....	120
Figure 4.5 : Quelques images de visages et d'oreilles ainsi que leurs codes LBP correspondants.....	120
Figure 4. 6 : Exemple de voisinages (4,1), (8,1), (16,2), et (8,2) circulaires.	122
Figure 4. 7 : Effet de rotation de l'image sur les points d'échantillonnage dans un voisinage circulaire.....	123
Figure 4. 8 : Exemple de la représentation d'une modalité biométrique (visage) à base de LBP.....	124
Figure 4. 9 : Ensemble de voisins linéairement symétrique pour différentes valeur de m et W	126
Figure 4. 10 : Application du 1DLBP sur une image d'oreille normalisée (masque de 08 voisins).	126
Figure 4.11 : Implémentation basique du LPQ: (a) Exemple de fréquence utilisée (b) Exemple de calcul par le descripteur.....	129
Figure 4. 12 : Caractéristiques statistiques et binarisées de l'image (BSIF: Binarized Statistical Image Features)	129
Figure 4. 13 : Banque de filtres BSIF de taille 7×7 et quantification de $8bits$ appris par des images naturelles.	131
Figure 4.14 : Quelques images de visages et d'oreilles ainsi que leurs codes BSIF (avec $l = 9$ et $n = 8$) correspondants.....	131
Figure 5.1 : Stratégie de la décomposition de l'image en plusieurs sous-blocs (décomposition horizontale-verticale).....	136
Figure 5. 2 : Quelques échantillons de la base d'images ORL.....	137
Figure 5. 3 : Quelques échantillons de la base d'images AR.	138
Figure 5. 4 : Quelques échantillons de la base d'images IIT Delhi.	139

Figure 5. 5 : Quelques échantillons normalisés de la base d'images IIT Delhi	139
Figure 5. 6 : Quelques échantillons de la base d'images USTB.	140
Figure 5.7 : Taux d'identifications au rang-1 en utilisant les paramètres optimaux de l'opérateur BSIF (filtre de taille 15×15 pixels et 12 bits).	149
Figure 5. 8 : Différence de la valeur de profondeur entre les balayages neutres et d'autres six expressions pour une personne donnée.....	151
Figure 5. 9 : Exemple d'application du masque sur des échantillons de la base d'images AR, pour séparer entre les régions rigides et semi-rigides: (a) images par défaut et (b) images avec masque.....	151
Figure 5. 10 : Exemple de séparation entre les régions rigides et semi-rigides.....	152
Figure A. 1 : Illustration de la méthode des k -plus proches voisins (k -NN) avec $k=5$	189
Figure A. 2 : Illustration de la marge maximale et les vecteurs supports.	190
Figure A. 3 : Projection des données non linéairement séparables dans un espace à grande dimension.	191

Introduction Générale

La biométrie est la science qui permet de reconnaître l'identité d'une personne sur la base de ses caractéristiques physiologiques, chimiques ou comportementales, telles que: le visage, l'iris, l'odeur, la façon de marcher ou la signature électronique...etc. Avec la nécessité de techniques solides de reconnaissance humaine dans les applications critiques, telles que: le contrôle d'accès sécurisé, le passage des frontières internationales et les applications légales, la biométrie se positionne comme une technologie viable qui peut être intégrée dans les systèmes de management d'identité à grande échelle. Les systèmes biométriques fonctionnent en vertu du principe que la plupart des caractéristiques biologiques de l'être humain soient distinctives pour chaque individu, puissent être acquises d'une manière fiable à l'aide des capteurs convenables et peuvent être représentées dans un format numérique. Ainsi, ces systèmes peuvent être considérés comme des moteurs de reconnaissance des formes et peuvent être incorporés dans divers marchés.

Plusieurs caractéristiques humaines ont été étudiées et testées. Ces modalités peuvent être encore subdivisées en différentes sous-catégories selon leur position respective dans le corps humain, telles que (i) attributs de la région de la main (ex., empreinte digitale, géométrie de la main,...), (ii) attributs de la région faciale (ex., visage et oreille), (iii) attributs de la région oculaire (ex., iris et rétine), (iv) attributs comportementaux (ex., façon de marcher, signature électronique,...) et (v) attributs médicaux-chimiques (ex., os, odeur, ADN,...). Bien que l'ADN, l'iris et l'empreinte digitale soient considérés parmi les modalités extrêmement fiables, elles reposent, malheureusement, sur la coopération du participant-capteur.

Dans le cadre de cette thèse, nous nous sommes concentrés par les attributs faciaux, à savoir: le *visage* et l'*oreille*, puisqu'ils sont souvent efficaces et ne nécessitent aucune coopération ou connaissance à priori de l'utilisateur. Par exemple, l'identification des criminels et des suspects dans les aéroports et dans les autres secteurs publics par des vidéos surveillances, est un scénario typique dans lequel apparaît l'importance des attributs faciaux. En outre, le mode de reconnaissance biométrique est plus accepté par l'homme, en raison de son caractère naturel.

La recherche dans la reconnaissance du visage est motivée non seulement par les défis fondamentaux de ce problème mais aussi par les nombreuses applications pratiques où l'identification humaine est nécessaire. La reconnaissance du visage, comme l'une des technologies biométriques pilotes, est devenue de plus en plus importante en raison des progrès rapides dans les technologies comme: les appareils photos numériques, l'internet et les téléphones intelligents, ainsi que l'augmentation des exigences de sécurité. La reconnaissance par visage a plusieurs avantages par rapport aux autres modalités biométriques: elle est naturelle, non-intrusive et facile à utiliser. Cependant, les conditions d'acquisition, telles que: la position du visage par rapport à la caméra, l'éclairage, les expressions faciales et la résolution de l'image du visage, ainsi que le vieillissement naturel causent de nombreux changements sur l'apparence du visage, celles-ci affectent négativement sur les performances de la reconnaissance. Des variations additionnelles peuvent être provoquées par des déguisements, des occlusions par lunettes, chapeaux et des cheveux, ainsi que les attaques malveillantes (ex., *Spoofing Attacks*) ajoutent aussi une autre complexité sur le visage. Le défi de la reconnaissance du visage consiste, donc, à identifier une personne sous la présence de toutes ces variations.

D'autre part, l'oreille humaine est considérée comme une nouvelle classe viable de la biométrie avec certains avantages additionnels par rapport au visage. En effet, l'oreille est riche en termes de caractéristiques; c'est une structure stable qui ne change pas considérablement avec l'âge et sa forme ne varie pas avec les expressions faciales, elle peut être aussi capturée à distance et sans aucune coopération de l'utilisateur, mais parfois elle peut être cachée par des cheveux, des écouteurs, une écharpe ou des boucles. Quoique les chercheurs ont développé plusieurs techniques, au cours de ces dernières années, qui utilisent des images d'intensités 2D, malheureusement, leurs performances se dégradent considérablement avec les variations de la pose et les conditions de l'image.

Un système de reconnaissance biométrique, en général, peut être divisé en deux activités fondamentales: *extraction des caractéristiques* et *classification*. Dans l'extraction des caractéristiques, il existe deux approches principales: *globale* et *locale*. Selon les études neuropsychologiques, le système visuel de l'être humain combine les caractéristiques globales et locales pour différencier et reconnaître les objets, et plus spécifiquement le visage. D'autre part, la plupart des travaux existants dans la littérature soient basés sur (i) des approches

globales (ex., l'analyse en composante principale (PCA)) ou (ii) des approches locales (ex., la correspondance par graphe de tas élastique (EBGM)). Contrairement aux travaux précédents, nous avons proposé, dans le cadre de cette thèse, d'utiliser une nouvelle approche d'extraction des caractéristiques basée principalement sur les descripteurs de texture locaux dans le but d'identifier les individus à partir des images d'intensités 2D de visages ou oreilles, dans des conditions non-contrôlées. Nous avons testé et considéré, en particulier, quatre descripteurs très récents qui sont: LBP (*Local Binary Pattern*), 1DLBP (*One Dimensional Local Binary Pattern*), LPQ (*Local Phase Quantization*), et BSIF (*Binarized Statistical Image Features*). La méthode proposée a la capacité de dériver facilement des modèles de caractéristiques efficaces qui combinent dans un vecteur de données unique la forme globale de l'objet analysé et la texture locale de son apparence; elle imite, d'une façon similaire, les capacités de l'être humain en reconnaissance d'objets ou visages, par sa particularité en codification des composantes à basses et hautes fréquences qui contribuent à la description locale et globale respectivement. Pour améliorer les taux d'identifications, nous avons effectué des efforts additionnels, tels que: des recherches dans les domaines de la neuroscience, la psychologie et l'embryologie pour trouver les composantes stables et plus importantes pour la reconnaissance, une recherche pour trouver les paramètres optimaux de chaque descripteur ainsi que l'utilisation des informations rigides/semi-rigides de chaque modalité.

Afin d'évaluer les performances de l'approches proposée, nous avons effectué des études sur la sensibilité des résultats par rapport aux rotations, éclairages, expressions faciales, occlusions et les régions rigides/semi-rigides. Nos tests ont été établis en utilisant des images fournies par les bases de données ORL, AR, IIT Delhi et USTB, les deux premières comportant des images de visages tandis que les deux dernières comportant des images d'oreilles, acquises chez des différentes personnes et sous différentes conditions.

Ce manuscrit est composé de cinq grands chapitres. Il est structuré comme suit :

Le premier chapitre « *Introduction à la biométrie et aux systèmes biométriques* » s'articule autour de deux parties. Dans la première partie, nous présentons des notions et des définitions de base liées à la biométrie. Nous évoquons ensuite le principe de fonctionnement des systèmes biométriques ainsi que les outils, généralement utilisés pour évaluer leurs performances. Nous finissons cette première partie par la présentation et une comparaison de quelques modalités biométriques les plus utilisées à nos jours. Dans la deuxième partie de ce

chapitre, nous accordons une attention particulière à la reconnaissance biométrique par *visages* et *oreilles* parmi les autres modalités biométriques, puisqu'elles constituent l'objectif de cette thèse. A la fin de cette partie, nous présentons « *la biométrie cachée* » en tant que nouvel axe de recherche en criminalistique et en sécurité biométrique, qui constitue un défi très important que nous voudrions l'exploiter dans des futurs travaux.

Intitulé « *Reconnaissance biométrique par visages* », nous présentons à travers ce deuxième chapitre les éléments essentiels pour la reconnaissance du visage par les humains et les machines, en prenant une perspective qui englobe des approches psychologiques, neurologiques et informatiques. Du côté humain, nous discutons les méthodes et les techniques utilisées en neurobiologie pour reconnaître les visages, le rôle de l'attention visuelle et la nature qui émerge les représentations. Du côté machine, nous présentons les technologies de reconnaissance de visages et les stratégies qu'elles utilisent afin de surmonter les défis pour un fonctionnement robuste sous différents paramètres de visualisation. Finalement, nous terminons le chapitre par la présentation de quelques défis ouverts, des solutions récentes proposées et les futures directions dans ce domaine.

Le troisième chapitre, intitulé « *Reconnaissance biométrique par oreilles* », est consacré à des études anatomiques et embryologiques de l'oreille humaine dans le but de connaître les emplacements sur lesquels nous pouvons atteindre une grande variation interindividuelle. Ensuite, nous récapitulons les approches de détection et reconnaissance automatique de l'oreille, dans le domaine de l'ingénierie. Finalement, nous présentons quelques perspectives sur les problèmes non-résolus et les futures recherches, spécialement dans le contexte de la surveillance intelligente et la criminalistique.

Après avoir introduit, dans les chapitres deux et trois, les principales méthodes d'extraction des caractéristiques en lien avec le visage et l'oreille, nous présentons dans ce quatrième chapitre, intitulé « *Caractérisation biométrique par descripteurs de texture locaux* », les techniques que nous avons utilisé et testé dans le but d'extraire des informations biométriques texturées. Nous étudions l'utilisation de quatre descripteurs de texture locaux très récents: LBP, 1DLBP, LPQ et BSIF.

Dans le cinquième chapitre « *Implémentation et études expérimentales* », nous testons et nous comparons les performances des descripteurs de texture locaux, cités précédemment, appliquées sur des données biométriques bidimensionnelles en utilisant des bases d'images de

visages et/ou oreilles en expérimentations. Nous analysons la sensibilité des techniques implémentées par rapport aux changements d'expressions faciales, de rotations et d'occlusions, en termes de précision de la classification et la complexité algorithmique

Nous achevons notre manuscrit par une « *Conclusion et perspectives* », dans laquelle, d'une part, nous concluons notre travail de trois années de recherche et d'autre part, nous établissons de manière étendue des perspectives, notamment en considérant l'implémentation d'un système biométrique multimodal impliquant le visage et l'oreille simultanément.

Chapitre 1

Introduction à la Biométrie et aux Systèmes Biométriques

1.1 Introduction

La biométrie est la science qui détermine l'identité d'un individu, elle se base sur des mesures physiologiques, chimiques ou comportementales d'un ou de plusieurs de ses attributs biologiques. La pertinence de la biométrie dans les sociétés modernes a été augmentée à cause du grand besoin de la sécurité et à la nécessité des systèmes de management (gestion) d'identités à grande échelle, qui s'appuient fonctionnellement sur la détermination précise de l'identité d'un individu, dans un contexte d'applications largement interconnectées. Comme exemples de ces applications: le partage des ressources informatiques dans un réseau public, l'accès de haute sécurité aux zones nucléaires, les transactions bancaires à distance, ou l'embarquement des vols commerciaux. En plus, la prolifération des services web (ex., les banques en ligne) et le déploiement des centres de services clientèles décentralisés (ex., les cartes de crédit) ont souligné la nécessité des systèmes de managements d'identité fiables pouvant accueillir un grand nombre d'individus.

Dans ce chapitre, nous introduisons tout d'abord quelques notions et définitions de bases liées à la biométrie, nous décrivons le principe de fonctionnement d'un système biométrique ainsi que les outils d'évaluations utilisés pour mesurer leurs performances, nous donnons un bref aperçu des modalités biométriques les plus répandues, tout en accordant une attention particulière à la reconnaissance par *visages* et *oreilles* parmi les autres modalités biométriques, puisqu'elles constituent l'objectif de cette thèse, nous présentons à la fin *la biométrie cachée* comme un nouvel axe de recherche émergeant dans le domaine de la sécurité des individus, illustré par quelques exemples et travaux réalisés.

1.2 Généralités et notions de bases en biométrie

La tâche primordiale d'un système de management d'identité est la détermination (ou la vérification) de l'identité d'un individu (ou d'une identité proclamée). Une telle action est très

importante pour une variété de raisons; mais dans la plupart des applications, le défi majeur est d'empêcher l'accès des imposteurs aux ressources protégées. Les méthodes traditionnelles d'authentification biométriques sont basées soit sur *une connaissance à priori* de la personne (ex., un mot de passe ou un code d'activation) ou sur *la possession* d'un objet (ex., une pièce d'identité, un badge ou une clef). Cependant, ce type de présentation d'identité peut être facilement perdu, partagé, oublié par son utilisateur, volé ou deviné par d'autres personnes. Aujourd'hui, *la biométrie* offre une solution naturelle et fiable, en aspect de management d'identité, par une utilisation entièrement automatisée ou semi-automatisée afin de reconnaître des individus en fonction de leurs caractéristiques biologiques [01]. A l'introduction de la biométrie, il est possible d'établir une identité qui répond à une question de type «*qui es-tu ?*» au lieu de «*qu'est ce que tu possèdes ?*» (Le cas d'une pièce d'identité) ou «*qu'est ce que tu rappelles ?*» (Le cas d'un mot de passe). Certaines applications utilisent la biométrie comme un outil supplémentaire avec les pièces d'identité ou avec les mots de passe donnant ainsi un niveau additionnel de sécurité. Ce type de renforcement est souvent appelé un *schéma d'authentification à double facteur*.

Dans un système de management d'identité particulier, l'*identité* d'un individu est considérée comme une information associée à cet individu [02]. Par exemple, une carte de crédit bancaire associé typiquement à un client, son nom, un mot de passe, son numéro de sécurité sociale, son adresse et sa date de naissance. Ainsi, l'identité du client dans cette application sera définie par ses attributs personnels (ex., nom, adresse, etc.). Le terme *authentification biométrique* est mieux approprié que *biométrie*, puisque ce dernier a été historiquement utilisé dans le domaine de statistique et d'analyse de données biologiques (terme particulièrement médical) [03].

L'efficacité d'un authentificateur (biométrique ou non-biométrique) est basée sur sa pertinence à une application particulière, ainsi que sur sa robustesse à divers types d'attaques malveillantes. *O'Gorman* [04] a dressé une liste des attaques qui peuvent être lancées contre les systèmes basés sur les mots de passes ou les possessions: (a) attaque du client (ex., deviner un mot de passe ou voler une pièce d'identité); (b) attaque du serveur (ex., l'accès à des fichiers, hébergés sur un serveur, contenant des mots de passes); (c) écoute clandestine (ex., écouter délibérément une conversation pour deviner un mot de passe); (d) répudiation (ex., proclamer qu'une possession est mal placée); (e) attaque du *cheval de Troie* (*Trojan*

Horse) (ex., placer un faux écran d'authentification pour voler les mots de passes); et (f) dénégalion du service (ex., désactiver le système délibérément en fournissant, à plusieurs fois, des mots de passes incorrects). Bien que plusieurs types de ces attaques puissent être interrompus en intégrant des mécanismes de défense appropriés, il n'est pas possible de gérer toutes les limites associées aux possessions et aux mots de passe.

La biométrie offre certains avantages tels que la *reconnaissance négative* et la *non-répudiation* qui ne peuvent pas être fournis par les mots de passe et les possessions [05]. La reconnaissance négative est le processus où un système détermine qu'un certain individu est inscrit dans un système tandis que cet individu refuse son existence. Ceci est très important dans les applications du déboursement d'assistance sociale où un imposteur tente de déclarer des bénéfices multiples sous différents noms. La non-répudiation est une manière de garantir qu'un individu qui a accédé à un certain service ne peut pas, plus tard, contredire son accès (ex., un utilisateur accède à certaines ressources informatiques et proclame, plus tard, qu'un imposteur a utilisé cette ressource avec des informations falsifiées).

Afin de déterminer l'identité d'un individu, les systèmes biométriques utilisent une variété de caractéristiques physiologiques, chimiques ou comportementales (voir Figure 1.1) qui sont: l'empreinte digitale, le visage, la géométrie de la main / du doigt, l'iris, la rétine, la signature électronique, la façon de marcher, la dynamique de la frappe sur un clavier, l'empreinte palmaire, la voix, l'oreille, la veine de la main, l'odeur ou l'information apportée par l'ADN [06]. Dans la littérature de la biométrie, ces caractéristiques biométriques sont préférées comme: *traits*, *indicateurs*, *identificateurs* ou *modalités*. Comme les systèmes biométriques ont leurs propres limites [07], ils ont plusieurs avantages par rapport aux systèmes de sécurité traditionnels, ils ne peuvent pas être volés ou partagés facilement.

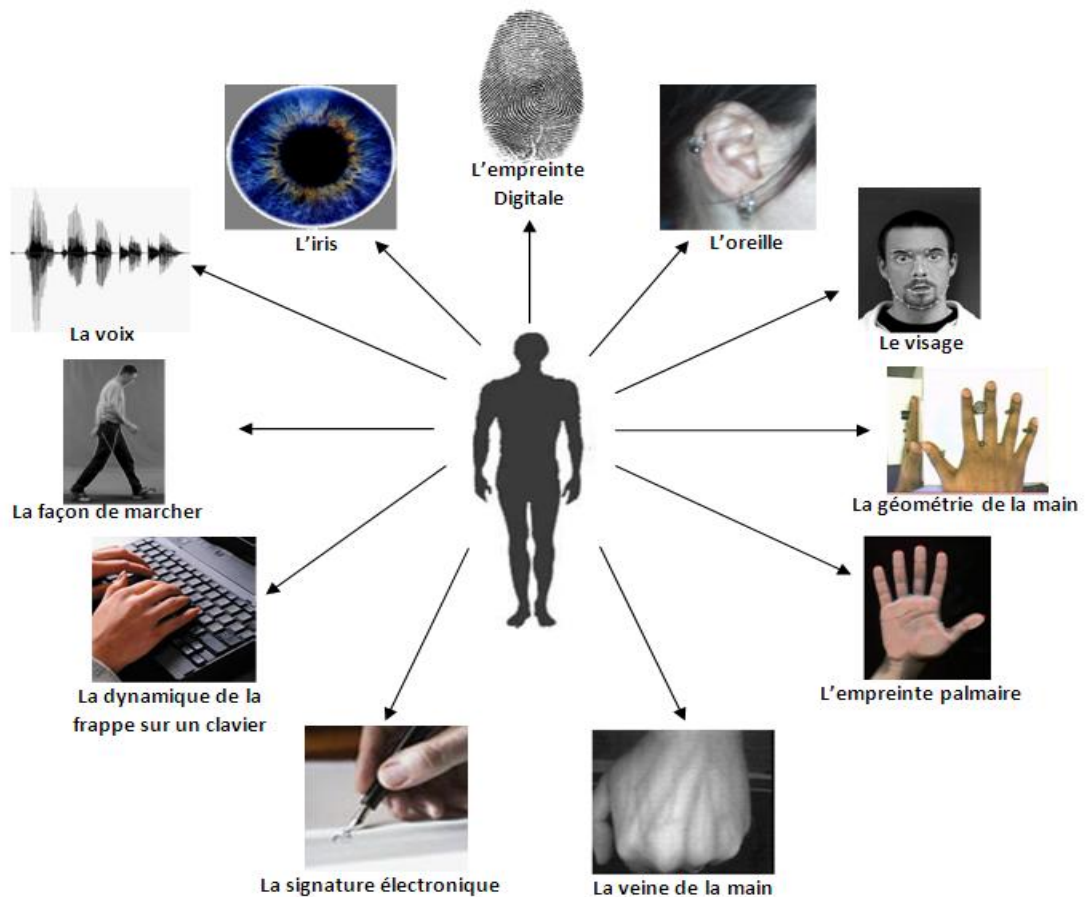


Figure 1. 1 : Exemples de modalités biométriques (physiologiques et comportementales) qui peuvent être utilisées pour authentifier un individu.

1.3 Architecture fonctionnelle d'un système biométrique

Un système biométrique est, généralement, composé de deux principaux modes (voir la Figure 1.2) qui sont le mode d'*apprentissage* appelé également mode d'enregistrement ou d'enrôlement et le mode de *reconnaissance* (vérification ou identification). Le premier mode consiste à enregistrer dans une base de données les caractéristiques physiques ou comportementales d'un individu sous forme d'un "*Modèle*" biométrique appelé aussi "*Template*" ou "*Signature*". Le deuxième mode consiste à tester les mêmes caractéristiques et à les comparer avec les modèles biométriques stockés dans la base de données. Si les données testées correspondent à un modèle biométrique enrôlé, l'individu est donc considéré comme reconnu [08].

1.3.1 Fonctionnement interne d'un système biométrique

Un système biométrique est essentiellement considéré comme un système de reconnaissance des formes [09] qui acquiert des données biométriques d'un individu, extrait des caractéristiques saillantes à partir des données biométriques, compare ces caractéristiques par rapport aux modèles stockés dans la base de données et exécute une action basée sur le résultat de la comparaison. Par conséquent, un système biométrique générique peut être vu comme un processus à quatre modules: un module de capture, un module de prétraitement et d'extraction des caractéristiques, un module de correspondance et un module de gestion de la base de données. Le fonctionnement de chaque module est détaillé ci-dessous [09]:

- a. **Module de capture:** Un lecteur ou un scanner biométrique approprié est nécessaire pour acquérir des données biométriques d'un individu. Par exemple, afin d'obtenir des images d'empreintes digitales, un capteur optique d'empreintes digitales est utilisé pour numériser les structures de frottement en arêtes du doigt. Le module de capture, définissant l'interface d'interaction homme-machine, joue un rôle très important en performances de chaque système biométrique. En effet, une mauvaise interaction homme-machine peut *engendrer automatiquement* une forte *augmentation* du *taux de fausses acceptations* (voir section 1.4) et, par conséquent, une faible acceptabilité par les utilisateurs. La capture des données biométriques peut être supervisée par un humain ou non en fonction de chaque application. Comme la plupart des modalités biométriques sont acquises sous forme d'images (à l'exception de la voix et de l'odeur qui sont respectivement basées sur l'audio et les produits chimiques), la qualité des données biométriques brutes peut être aussi affectée par des bruits liés aux caractéristiques technologiques de la caméra utilisée.
- b. **Module de prétraitement & d'extraction des caractéristiques:** La qualité des données biométriques acquises par le capteur biométrique est d'abord évaluée afin de déterminer sa pertinence pour un traitement ultérieur. Typiquement, les données acquises doivent être passées par un algorithme de restauration qui améliore leur qualité. Cependant, dans certains cas où la qualité des données biométriques est mauvaise, l'utilisateur est invité à présenter de nouveau ses caractères biométriques, qui seront traitées, et à la fin, un ensemble de caractéristiques saillantes et discriminatives sera extrait pour représenter / modéliser le trait biométrique. Durant la phase d'apprentissage, cet ensemble (vecteur)

de caractéristiques est stocké dans la base de données qui est généralement appelé *modèle* ou *gabarit*.

- c. Module de correspondance:** Le vecteur des caractéristiques, extrait, est comparé avec les modèles stockés dans la base de données pour générer des scores de correspondances. Le module de correspondance encapsule également un module de prise de décision, dans lequel les scores de correspondances sont utilisés soit pour valider une identité proclamée ou pour fournir un classement d'identités, inscrites dans la base de données, afin d'identifier cet individu.
- d. Module de gestion de la base de données:** La base de données agit comme un référentiel de données biométriques. Durant la phase d'apprentissage, le vecteur des caractéristiques biométriques, extrait, à partir d'un échantillon biométrique brut (modèle) est stocké dans la base de données, éventuellement, avec quelques informations biographiques caractérisant cet utilisateur (ex., nom, numéro d'identification personnelle (PIN), adresse, etc.).

Le *modèle* d'un utilisateur peut être extrait à partir d'un seul échantillon biométrique, ou généré par un traitement de plusieurs échantillons. Par exemple, le modèle de minuties d'une empreinte digitale est extrait après la création d'une mosaïque de multiples échantillons de la même empreinte. Certains systèmes biométriques enregistrent plusieurs modèles d'un même individu dans le but de diminuer les variations intra-classes (voir section 1.4) associées à chaque individu. Un système de reconnaissance par visage, comme exemple, peut utiliser plusieurs modèles d'un même individu où chaque modèle correspond à une position faciale différente. Selon l'application utilisée, le modèle biométrique peut être enregistré dans la base de données du système biométrique ou dans un support numérique approprié (ex., carte magnétique).

Dans la littérature de la reconnaissance par visages, les images biométriques brutes stockées dans la base de données sont souvent nommées "*galerie-images*" tandis que celles acquises durant l'authentification sont nommées "*probe-images*" qui signifient respectivement "*images enregistrées*" et "*requêtes*" ou "*images d'entrées*".

1.3.2 Vérification et Identification

Selon le contexte de chaque application, un système biométrique peut fonctionner en mode vérification ou en mode identification (voir Figure 1.2). En mode vérification [10], le système évalue l'identité d'une personne en comparant les données biométriques capturées avec son (ses) propre(s) modèle(s) enregistré(s) dans la base de données du système. Dans ce type d'application, un individu qui veut être reconnu par le système doit proclamer son identité, habituellement, par son numéro d'identification personnelle (PIN), par son nom d'utilisateur ou par sa carte magnétique, le système effectuera une comparaison de type *un-contre-un* pour déterminer si cette proclamation est vraie ou fausse (Ex., «*Est-ce que ces données biométriques appartiennent à Ahmed ?*»). La vérification est typiquement utilisée pour une reconnaissance positive afin d'empêcher l'utilisation d'une même identité par plusieurs personnes.

En mode identification [10], le système identifie un individu en recherchant le modèle enrôlé qui représente la meilleure correspondance parmi tous les modèles d'utilisateurs stockés dans la base de données. Par conséquent, le système effectue une comparaison de type *un-contre-tous* afin d'établir l'identité de cet individu (ou par un échec si cet individu n'existe pas dans la base de données du système) sans faire de proclamation préalable d'identité (Ex., «*A quelle personne appartiennent ces données biométriques ?*»). L'identification est une composante essentielle pour les applications de reconnaissances négatives; l'objectif de ce type de reconnaissance est d'empêcher l'utilisation de plusieurs identités par un seul individu. Comme les méthodes traditionnelles de reconnaissances d'individus telles que les mots de passe et les possessions peuvent fonctionner en mode positif, le mode négatif peut être uniquement établi à l'aide de la biométrie.

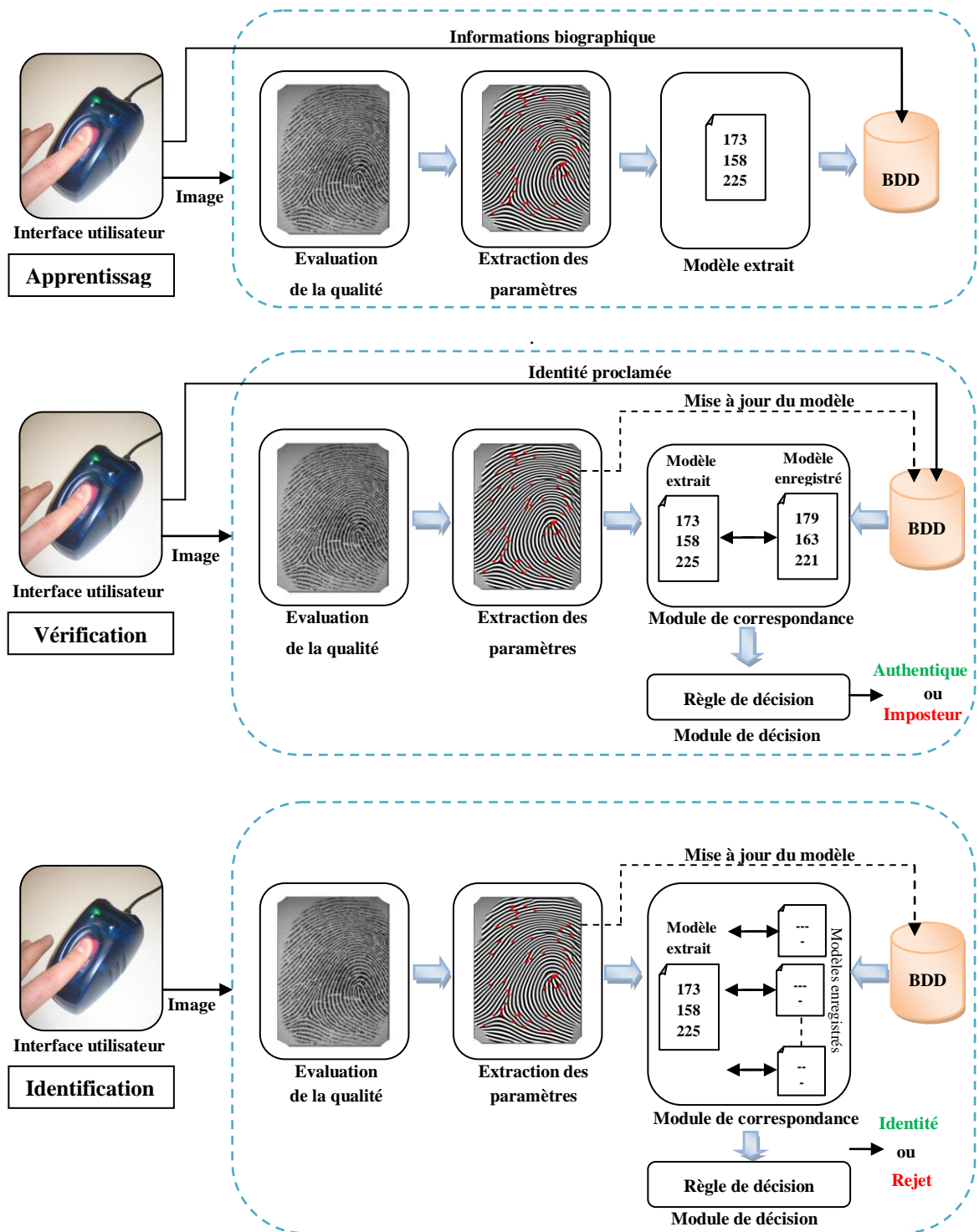


Figure 1. 2 : Architecture fonctionnelle d'un système biométrique. Diagramme des processus d'apprentissage, vérification et identification.

1.4 Performances d'un système biométrique

Dans les systèmes basés sur les mots de passe, une correspondance *parfaite* est *nécessaire* entre les deux chaînes de caractères pour valider l'identité d'un individu. Par contre, les systèmes biométriques rencontrent rarement deux modèles biométriques d'un même utilisateur présentant exactement les mêmes vecteurs de caractéristiques, en raison de: mauvaises conditions (ex., une empreinte digitale qui contient un bruit lié à un défaut du capteur), changements des caractéristiques biométriques de l'utilisateur (ex., une maladie respiratoire effectuant la reconnaissance du speaker), changements des conditions ambiantes (ex., le changement du niveau d'illumination en reconnaissance par visage) et variations en interaction utilisateur-capteur (ex., iris occlue ou empreinte digitale partielle). Il est donc rare d'avoir deux modèles biométriques exactement similaires provenant du même utilisateur. En effet, une correspondance parfaite entre deux vecteurs de caractéristiques peut indiquer la possibilité qu'il y ait une attaque malveillante lancée contre le système. La variabilité observée entre les vecteurs de caractéristiques provenant d'un même individu est appelée variation "*intra-classes*", et la variabilité observée entre les vecteurs de caractéristiques provenant de différents individus est appelée une variation "*inter-classes*". Un vecteur de caractéristiques *utile* présente une petite variation *intra-classe* et une grande variation *inter-classes*.

1.4.1 Test de vérification

Dans le cas de vérification, un utilisateur doit présenter son trait biométrique et proclamer son identité (ex., «*je suis Mohamed*») au système biométrique, ce dernier calcul un score / une distance de similarité (ou de dis-similarité) (comprise entre 0 et 1) entre le vecteur de caractéristiques extrait à partir du trait présenté et le vecteur de caractéristiques du même utilisateur enregistré dans la base de données (vecteur enrôlé). Plus le score de similarité est proche de 0, plus le système est confiant que les deux modèles proviennent d'un même individu, dans ce cas l'individu est nommé: *genuine / authentique*. Plus le score est proche de 1, plus le système est sûr que les modèles proviennent de deux individus différents, dans ce cas l'individu proclamé est nommé: *imposteur*. La décision est prise par le choix d'un seuil de décision θ ; un score de similarité inférieur ou égale à θ indique que les deux modèles comparés appartiennent à la même personne, et un score supérieur à θ indique que les deux

modèles comparés proviennent de deux personnes différentes [11]. Le test de vérification est formulé mathématiquement de la manière suivante:

- Soit X_Q le vecteur de caractéristiques de la personne proclamée I .
- Soit X_t le vecteur de caractéristiques de la personne I stocké dans la base de données.
- Soit $S(X_Q, X_t)$ la fonction de similarité entre les deux vecteurs X_Q et X_t ; la fonction S retourne un score de similarité entre les mesures biométriques de la personne enrôlée et la personne proclamée.
- Le test de vérification est défini par la fonction (I, X_Q) tel que :

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

- Où w_1 indique que la personne proclamée est *authentique* et w_2 qu'elle est *imposteur*.

L'évaluation d'un système biométrique en mode vérification nécessite d'établir une distribution des scores de similarités des utilisateurs authentiques par des comparaisons intra-classes, ainsi, d'établir une distribution des scores de similarités des utilisateurs imposteurs par des comparaisons inter-classes. Avec ces deux distributions, nous distinguons deux types d'erreurs de vérification liés à des mauvaises décisions [12]:

- **Le Taux de Faux Rejets (*False-Rejection Rate*), noté **FRR**** : il exprime le pourcentage des utilisateurs authentiques faussement rejetés par un système biométrique. En d'autres termes, ce taux représente les données de test authentiques qui ont été incorrectement rejetées et considérées comme des imposteurs. FRR est aussi nommé: *False Non-Match Rate* (FNMR).
- **Le Taux de Fausses Acceptations (*False-Acceptance Rate*), noté **FAR**** : il exprime le pourcentage des utilisateurs imposteurs faussement acceptés par un système biométrique. En d'autres termes, il représente les données de test imposteurs qui ont été incorrectement acceptées et considérées comme authentiques.

En pratique, la valeur du FAR est affectée par le nombre de fausses acceptations (FA) divisé par le nombre de tests imposteurs dans la base de données (N). Tandis que, la valeur du FRR est affectée par le nombre de faux rejets (FR) divisé par le nombre de tests authentiques (M). En plus, les deux taux FAR et FRR sont dépendants du seuil de décision θ fixé dans le

module de décision. En effet, en faisant varier la valeur de θ , les deux taux d'erreurs seront modifiés d'une manière importante. Le calcul du FAR et FRR en fonction de θ est donné par :

$$FAR(\theta) = \frac{FA(\theta)}{N} \quad (1.2)$$

$$FRR(\theta) = \frac{FR(\theta)}{M} \quad (1.3)$$

Le choix optimal de la valeur du seuil de décision θ est très important puisqu'il influence directement sur la performance et la fiabilité du système biométrique. En effet, une valeur élevée du seuil θ entraîne l'apparition d'un grand nombre de faux rejets, par contre d'une faible valeur du seuil, résulte un nombre important de fausses acceptations [13]. La valeur la plus optimale du seuil θ pour faire un équilibre entre le FRR et le FAR correspond à l'endroit où le FAR = FRR est noté par le *point d'équivalence des erreurs (Equal Error Rate (EER))*; ce dernier est déterminé par le point d'intersection entre la courbe des taux de fausses acceptations et la courbe des taux des faux rejets. Un exemple de détermination du point EER est illustré dans la figure 1.3.

Les deux taux FRR et FAR en différentes valeurs du seuil θ peuvent être récapitulés en utilisant la courbe : *Receiver Operating Characteristic (ROC)* [14]. Cette courbe trace le taux de faux rejets en fonction du taux de fausses acceptations. Elle sera tracée d'une manière paramétrique en fonction des valeurs du seuil θ . Un exemple de détermination du point EER en utilisant la courbe ROC est illustré dans la figure 1.4. Plus cette courbe tend à rapprocher la forme du repère, plus le système est performant, c'est-à-dire qu'il possède un taux de reconnaissance global élevé.

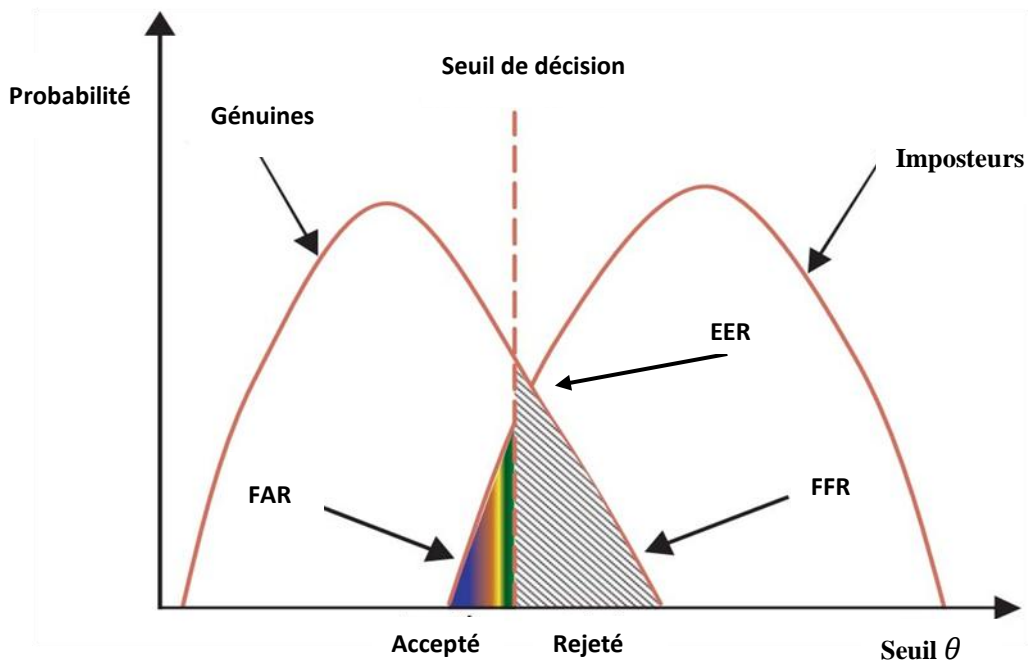


Figure 1.3 : Distribution des scores de similarités des imposteurs et authentiques en mode vérification. Illustration de détermination du point d'équivalence des erreurs dans un système biométrique.

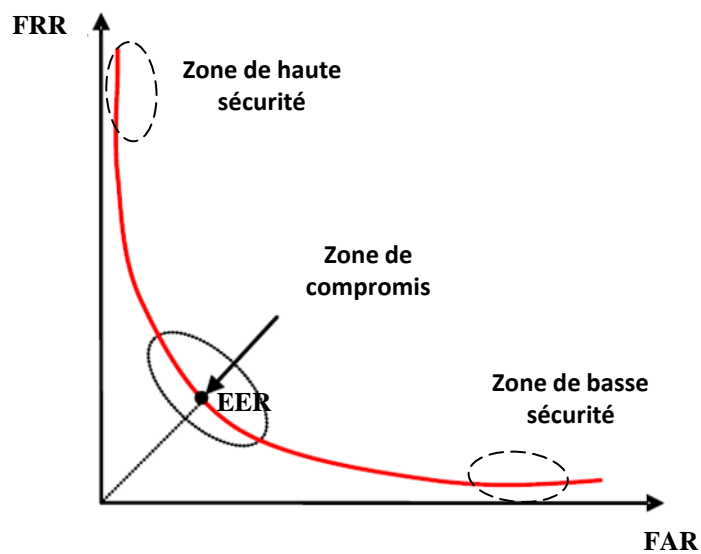


Figure 1.4 : Courbe ROC « Receiver Operating Characteristic ».

1.4.2 Test d'identification

Dans le cas d'identification, le vecteur des caractéristiques en entrée est comparé avec tous les modèles enregistrés dans la base de données afin de déterminer le meilleur modèle de correspondance (i.e., la meilleure correspondance). Ce dernier est déterminé en examinant les scores de correspondances qui concordent à toutes les comparaisons possibles, en affectant l'identité au modèle qui correspond au plus petit score de similarité. *Le taux d'identification (Identification Rate)*, noté IR, exprime le pourcentage des données (personnes) de tests qui ont été correctement identifiées (CI) par rapport au nombre total de données utilisées en tests (N). Le test d'identification est formulé mathématiquement de la manière suivante :

- Soient X_Q l'ensemble des vecteurs de caractéristiques inscrits dans la base de données.
- Soit $I_K, k \in \{1, 2, \dots, N, N + 1\}$ un vecteur de caractéristiques utilisé pour le test.
- Pour déterminer l'identité de I_K , nous calculons la fonction (I, X_Q) définie par :

$$(I, X_Q) = \begin{cases} I_K & \text{Si } \min_k \{S(X_Q, X_{1_K})\} \leq \theta, \\ I_{N+1} & \text{Sinon} \end{cases} \quad K = 1, 2, \dots, N \quad (1.4)$$

- Où I_1, \dots, I_N sont des identités enrôlées, I_{N+1} une identité rejetée, X_{1_K} la signature biométrique (meilleure) qui correspond à l'identité I_K , et θ est le seuil de décision.
- Le taux d'identification est formulé par :

$$IR(\theta) = \frac{CI(\theta) \times 100}{N} \quad (1.5)$$

Le taux d'identification représente la mesure la plus couramment utilisée, mais il n'est pas toujours suffisant. Cependant, des questions et des alternatives ont été supposées telles que: «*Est-ce que le meilleur modèle correspond vraiment à l'identité correcte ?*» ou «*Est-ce que l'un des k premiers meilleurs modèles correspond à l'identité correcte ?*». Le *taux d'identification au rang- k (IR_k)* indique le pourcentage des données (personnes) de tests qui ont été correctement identifiées parmi les k premiers top modèles de correspondances. La performance du *rang- k* peut être récapitulée en utilisant la courbe: *Cumulative Match Characteristics (CMC)* [15] qui trace les IR_k en fonction de k où $k = 1, 2, \dots, M$ et M représente le nombre de personnes enrôlées; cette courbe représente une probabilité que le bon choix se trouve parmi les meilleures k premières classes comme montré dans la figure suivante :

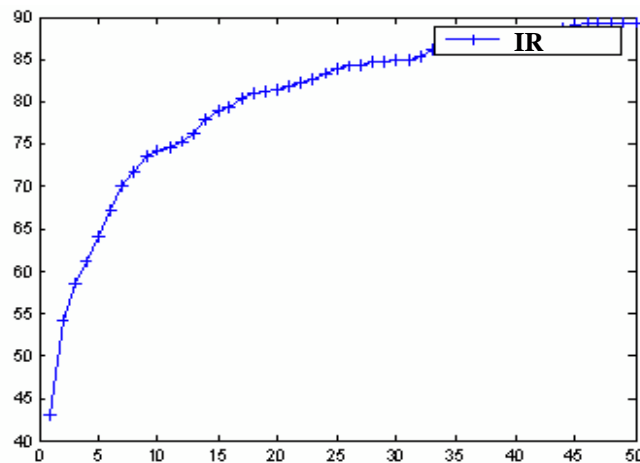


Figure 1.5 : Courbe CMC.

Les performances d'un système biométrique pour une application donnée sont principalement basées sur les taux d'erreurs et les taux d'échecs discutés ci-dessus. Autres facteurs tels que: le coût du système, l'acceptabilité par utilisateur, la facilité d'utilisation, la fiabilité du capteur, etc. déterminent aussi la convenance d'un système biométrique pour une application particulière.

1.5 Applications de la biométrie

Dans notre société largement interconnectée, la détermination de l'identité d'un individu avec une précision élevée est devenue un défi critique. Des questions de type : « *Est-il vraiment ce qu'il proclame être ?* », « *Est-ce qu'il est autorisé à utiliser ce service ?* » ou « *Est-il dans la liste noire signalée par le gouvernement ?* », sont habituellement posées dans divers scénarios. La nécessité des techniques d'authentification fiables est augmentée suite aux préoccupations croissantes qui concernent la sécurité et les progrès rapides en communication, réseau et mobilité. Ainsi, la biométrie est de plus en plus intégrée dans diverses applications ou bien pouvant être classées en trois groupes principaux qui sont:

- a. Applications commerciales :** telles que l'accès à un réseau d'ordinateurs, la sécurité des données électroniques, le e-commerce, l'accès d'internet, l'utilisation des cartes de crédit bancaire, le contrôle d'accès physique, mobile phone, la gestion des registres médicaux ou l'apprentissage à distance, etc.

- b. Applications gouvernementales :** telles que les cartes d'identité (ID cards), la sécurité sociale, le contrôle des frontières, le contrôle des passeports, le déboursement en assistance sociale ou en permis de conduite, etc.
- c. Applications légales :** telles que l'identification des corps humains, les enquêtes criminalistiques ou la détermination parentèle, etc.

1.6 Caractéristiques de la biométrie

Un certain nombre de caractéristiques sont utilisées pour faire une analyse comparative sur les performances de la modalité biométrique. Chaque modalité possède ses propres avantages et inconvénients, par contre, le choix d'une modalité biométrique pour une application donnée dépend d'une variété de paramètres liés à la nature et les exigences de l'application et aux propriétés de la modalité. *Jain et al.* [16] ont identifié sept paramètres afin de déterminer la convenance de la modalité physique ou comportementale, généralement employés dans une application biométrique, qui sont:

- a. Universalité:** Chaque individu, qui accède à une application, doit posséder le trait biométrique, c'est-à-dire que le trait biométrique existe chez tous les individus.
- b. Unicité:** Le trait biométrique doit être suffisamment différent par rapport aux autres individus dans une population donnée.
- c. Stabilité:** Le trait biométrique d'un individu doit être suffisamment stable et invariant au cours du temps. Un trait qui se change n'est pas préférable dans l'authentification biométrique.
- d. Mesurabilité:** Le trait biométrique doit être facilement mesurable et quantifiable à l'aide d'un capteur approprié qui ne cause aucun désagrément à l'individu. En plus, les données acquises doivent être favorables pour un prétraitement ultérieur afin d'extraire des informations saillantes.
- e. Performance:** Signifie que l'authentification doit être précise et rapide.
- f. Acceptabilité:** Indique que le trait biométrique utilisé doit être bien accepté par les utilisateurs du système.
- g. Non-reproductibilité / Circonvention:** Concerne la facilité ou non à falsifier un trait biométrique (ex., utilisation d'une fausse empreinte afin d'empêcher l'accès à un système).

Il n'existe aucun trait biométrique *exceptionnel* qui satisfait et répond efficacement à toutes les exigences (ex., précision, coût et praticabilité) imposées par les applications (ex., gestion des droits digitaux, contrôle d'accès ou gestion d'assistance sociale). En d'autres termes, il n'existe aucune modalité biométrique *idéale* mais certaines sont *admissibles*. La pertinence d'une modalité biométrique à une application spécifique dépend de la nature et des exigences de l'application ainsi que des propriétés de la modalité. Une brève introduction à certaines modalités biométriques communément utilisées est donnée ci-dessous:

- a. **Visage:** L'identification par visage est une méthode non-intrusive (cela veut dire qu'elle n'atteint pas l'intimité de l'individu), les attributs faciaux sont généralement les caractéristiques biométriques les plus communs et les plus utilisés par les humains pour différencier les individus. Les applications de reconnaissances faciales ont été évoluées d'une authentification statique à une authentification dynamique contrôlée, aujourd'hui certaines authentifications sont non contrôlées (i.e., avec des fonds encombrés). Les approches les plus connues dans la reconnaissance par visages sont basées [17] sur la localisation et la forme des attributs du visage, comme les yeux, les sourcils, le nez, les lèvres, le menton, et les relations spatiales entre eux (ex., distances et angles entre les attributs), ou sur l'analyse globale de l'image du visage, qui représente ce dernier comme une combinaison pondérée d'un certain nombre de visages canoniques. Bien que les performances des systèmes de reconnaissance par visages commercialisés sont raisonnables, elles imposent toujours un certain nombre de restrictions sur la façon d'obtenir les images faciales, elles nécessitent souvent un fond fixe et simple avec un éclairage contrôlé. Ces systèmes ont des difficultés en correspondance avec les images de visages capturés à partir de deux points de vue différents, dans des conditions d'éclairages différentes et à des moments distincts. Pour le fonctionnement d'un système de reconnaissance de visages dans la pratique, le système doit automatiquement [18] détecter si un visage est présent ou non dans l'image acquise, localiser le visage s'il existe et reconnaître le visage d'un point de vue général (i.e., de n'importe quelle position) et dans différentes conditions ambiantes.
- b. **Iris:** L'iris est la région annulaire de l'œil délimitée par la pupille et la sclérotique (le blanc de l'œil) de chaque côté. La texture visuelle de l'iris est formée au cours du développement fœtal et se stabilise au cours des deux premières années de la vie (la

pigmentation, cependant, continue le changement (l'évolution) pendant une longue durée). La texture complexe de l'iris comporte des informations très distinctives et utiles pour différencier et reconnaître les individus [19]. La précision et la rapidité des systèmes de reconnaissance actuels basés sur l'iris sont très intéressantes, en effet elles soutiennent la faisabilité de gérer des systèmes de management d'identité à grande échelle. Chaque iris est distinctif, même ceux des vrais *jumeaux* sont différents. Il est possible de détecter les lentilles de contacts imprimés avec des faux iris [20]. Le mouvement hippie de l'œil peut être utilisé aussi comme une mesure pour cette modalité biométrique. Bien que les premiers systèmes de reconnaissances biométriques basés sur l'iris ont été très chers et nécessitent une interaction utilisateur-capteur, les nouveaux systèmes sont devenus plus faciles à utiliser et moins coûteux [21].

- c. **Empreinte digitale:** L'être humain a utilisé les empreintes digitales, depuis plusieurs décennies, en criminalistique et en identification biométrique. Le taux de correspondance (identification) à l'aide d'empreintes digitales a été montré d'être très élevé [22]. Une empreinte digitale est le motif de crête et de vallées sur la surface au bout d'un doigt, dont la formation est établie aux cours des sept premiers mois du développement fœtal. Il a été déterminé, d'une manière empirique, que les empreintes digitales des vrais jumeaux sont différentes [23]. Aujourd'hui, la plupart des capteurs d'empreintes ne sont pas coûteux; le coût marginal d'intégration des données biométriques, à base d'empreintes digitales, dans les applications de sécurité (ex., un ordinateur portable) est devenu abordable. L'utilisation de multiples empreintes digitales de la même personne fournit des informations additionnelles permettant une identification à grande échelle, qui implique des millions d'identités. Le problème qui se pose avec ces systèmes, est la nécessité d'une énorme quantité de ressources informatiques (ex., espace mémoire), spécialement en mode identification. Finalement, les empreintes digitales d'une petite fraction de la population peuvent être inappropriées pour une identification automatique en raison de facteurs génétiques, de vieillissements, environnementaux ou professionnels (ex., les ouvriers peuvent avoir un certain nombre de coupures et de contusions sur leurs empreintes).
- d. **Géométrie de la main:** Les systèmes d'identification par géométrie de la main sont basés sur un certain nombre de mesures, qui sont prises à partir de la main humaine y compris sa forme, la taille de la palme, les longueurs et les largeurs des doigts [24,25].

Cette technique est très simple, relativement facile à utiliser et non coûteuse. Les facteurs environnementaux tels que le climat sec et les anomalies individuelles (ex., la peau sèche) peuvent affecter, négativement, la précision de ce type de systèmes. Cependant, la géométrie de la main n'est pas connue comme une modalité très distinctive, ainsi les systèmes de reconnaissances basés sur cette modalité ne peuvent pas être utilisés pour identifier un individu à partir d'une grande population. En outre, les informations de la géométrie de la main sont variantes durant la période de croissance des enfants. Ainsi, Les bijoux d'un individu (ex., les anneaux) ou des limitations dans la dextérité (ex., de l'arthrite) peuvent causer des problèmes lors de l'extraction de l'information géométrique de la main. La taille physique d'un système basé sur cette modalité est grande; elle ne peut pas être intégrée dans certains dispositifs comme les ordinateurs portables. Il existe aussi des systèmes d'authentification qui se basent uniquement sur la mesure de quelques doigts au lieu de la main entière; ces appareils sont plus petits que ceux utilisés pour la géométrie de la main.

- e. **Empreinte palmaire:** Les palmes de la main humaine contiennent des motifs de crêtes et de vallées qui sont peu similaires aux empreintes digitales. La surface de la palme est très grande que la surface d'un doigt et par conséquent, les empreintes palmaires sont plus distinctives que les empreintes digitales [26,27]. Cependant, leurs capteurs prennent une grande surface ce qui les rend plus volumineux et plus coûteux. Les palmes humaines contiennent aussi des caractéristiques additionnelles et distinctives telles que les lignes principales et les rides qui peuvent être capturées par un capteur de faible résolution. Lors de l'utilisation d'un capteur d'empreinte palmaire de haute résolution, toutes les caractéristiques de la main telles que la géométrie, les arêtes de vallées (ex., les minuties et les points singuliers (ex., le point Delta)), les lignes principales et les rides peuvent être combinées afin d'établir un système biométrique fortement précis.
- f. **Oreille:** Au cours de plusieurs années, l'oreille humaine a été utilisée comme un moyen d'identification en médecine légale. L'oreille humaine possède une richesse d'information qui se situe sur une surface 3D incurvée, cette richesse d'information a attiré l'attention des scientifiques légaux [28]. La morphologie extérieure de l'oreille est relativement stable durant une période de temps acceptable pour les applications biométriques. Plusieurs études sur la stabilité extérieure de l'oreille (i.e., l'auricule) suggèrent que la forme de l'oreille mûrit très tôt tandis que son expansion se poursuit

mais à un taux très faible [29]. Les images d'oreilles peuvent être acquises simultanément avec les images du visage et employées ensemble pour améliorer d'une manière significative la précision de la reconnaissance. Il est possible aussi d'employer l'oreille et le visage comme une pièce complémentaire d'information, spécialement dans les applications de surveillance, de suivi ou d'authentification personnelle continue [30]. Les avantages principaux liés à l'utilisation des images d'oreilles 2D en tant que modalité biométrique résident dans la stabilité contre les variations en expression faciale, l'immunité relative contre les soucis d'intimité et la convenance d'acquérir secrètement des images pour les applications de surveillance. Au fil de la *décennie passée*, les recherches ont légèrement évolué pour développer des technologies automatisées d'identification par oreille. Cependant, des efforts significatifs sont encore exigés pour améliorer la détection d'oreille, la segmentation et la possibilité d'identification dans le but de faire un déploiement dans la surveillance et dans les autres applications commerciales.

- g. Voix:** La voix est considérée comme une combinaison entre les caractéristiques biométriques physiques et comportementales. Les caractéristiques physiques de la voix d'un individu sont basées sur la forme et la taille des appendices (ex., les tractus vocaux, la bouche, les cavités nasales et les lèvres) qui sont utilisées dans la synthèse du son. Ces caractéristiques physiques de la parole humaine sont invariantes pour chaque individu, par contre, l'aspect comportemental de la parole se change au cours du temps en raison de l'âge, des conditions médicales (ex., rhume) et de l'état émotionnel. La voix n'a pas été connue comme une modalité très distinctive et n'est pas appropriée pour une identification à grande échelle. Un système de reconnaissance vocale de type *texte-dépendant* est basé sur l'expression d'une phrase fixe et prédéterminée. Par contre, un système de reconnaissance vocale de type *texte-indépendant* identifie un individu à la base de ce qu'il parle. L'implémentation des systèmes de type *texte-indépendant* est plus difficile par rapport aux systèmes de type *texte-dépendant*, mais elle offre plus de sécurité et protection contre les attaques malveillantes. L'inconvénient des systèmes de reconnaissance vocale est que les caractéristiques de la parole sont sensibles à certains facteurs comme le bruit [31]. La reconnaissance vocale est plus appropriée dans les applications qui se basent sur le téléphone malgré la dégradation de la qualité de la voix, typiquement, à travers le canal de transmission.

- h. Dynamique de la frappe sur un clavier:** Cette modalité n'est pas unique pour chaque individu, mais elle est utilisée afin d'offrir des informations suffisantes permettant de faire une vérification d'identité [32]. La dynamique de la frappe sur un clavier est une modalité comportementale; nous pouvons remarquer une grande variation intra-classe dans les modèles de la dynamique de frappe à cause des changements en état émotionnel, la position de l'utilisateur par rapport au clavier et le type du clavier utilisé, etc. Les frappes d'un individu peuvent être aussi surveillées discrètement. Avec cette modalité, nous pouvons faire *une vérification continue* de l'identité d'un individu, après son accès à une ressource par l'utilisation d'une modalité solide telle que: l'iris ou l'empreinte digitale.
- i. Signature électronique:** La manière avec laquelle un individu signe son nom est connue comme une caractéristique de cet individu. Bien que la signature exige un contact avec les instruments d'écriture (ex., stylo et papier) et un effort de la part de l'utilisateur, elle a été acceptée comme une méthode d'authentification légale par les gouvernements et dans les transactions commerciales. Avec la prolifération des ordinateurs et tablettes portables, la signature électronique a émergé comme un moyen d'authentification dans ces dispositifs. La signature est une modalité comportementale qui peut se changer en une certaine période du temps, elle est influencée par les conditions physiques et émotionnelles du signataire. Les signatures de certaines personnes se changent sensiblement, c'est-à-dire les mêmes impressions successives de leurs signatures sont très différentes. En plus, les faussaires professionnels sont capables de reproduire des signatures qui dupent les systèmes de vérification par signatures [33].
- j. Façon de marcher:** Elle se réfère à la manière dont une personne marche et c'est l'une des rares modalités biométriques qui peuvent être utilisées pour reconnaître des personnes à distance. Cette modalité comportementale est très appropriée dans les scénarios de surveillances où l'identité d'un individu peut être subrepticement déterminée. La plupart des algorithmes de reconnaissances par la façon de marcher cherchent à extraire une silhouette humaine dans le but de dériver des attributs spatio-temporels d'un individu en mouvement. Par conséquent, la sélection d'un bon modèle pour représenter le corps humain est très importante pour le fonctionnement efficace de ce type de systèmes. Quelques algorithmes utilisent le *flot optique* associé à un ensemble de points dynamiques du corps humain, extrait en mouvements, afin de décrire la marche

d'un individu. Les systèmes basés sur la marche offrent aussi la possibilité de suivre un individu pendant une longue durée. Cependant, la marche d'un individu est influencée par plusieurs facteurs y compris le choix des chaussures, la nature d'habillement, l'affliction des jambes ou la surface de marche, etc. [34]

D'autres modalités biométriques se développent actuellement telles que la biométrie par la géométrie des veines de la main [35,36] et la biométrie par la géométrie des veines du doigt [37,38]... Le tableau 1.1, compare certaines modalités biométriques en fonction de leurs propriétés. Malheureusement, il est très difficile d'avoir une modalité biométrique qui satisfait toutes les caractéristiques et exigences citées précédemment. Donc, le choix d'une modalité biométrique pour une application donnée dépend toujours des besoins et exigences liés à cette application.

Tableau 1. 1 : Comparaison entre les modalités biométriques [39].

Modalité	Universalité	Unicité	Performance	Stabilité	Acceptabilité	Mesurabilité	Circonvension
Visage	Elevée	Elevée	Moyenne	Elevée	Elevée	Elevée	Faible
Empreinte Digitale	Moyenne	Elevée	Moyenne	Moyenne	Moyenne	Elevée	Elevée
Oreille	Moyenne	Elevée	Elevée	Elevée	Elevée	Elevée	Faible
Iris	Moyenne	Elevée	Moyenne	Faible	Faible	Elevée	Faible
Empreinte Palmaire	Moyenne	Moyenne	Moyenne	Moyenne	Moyenne	Elevée	Moyenne
Signature	Faible	Elevée	Moyenne	Moyenne	Moyenne	Elevée	Elevée
Voix	Elevée	Moyenne	Moyenne	Elevée	Elevée	Elevée	Elevée
Façon de Marcher	Elevée	Moyenne	Moyenne	Elevée	Elevée	Moyenne	Faible
Frappe sur clavier	Moyenne	Moyenne	Moyenne	Moyenne	Elevée	Moyenne	Moyenne

Malgré leurs fiabilités, les systèmes d'identification biométrique uni-modaux ne garantissent pas la reconnaissance. La biométrie multimodale vient remédier à ce problème; elle consiste à utiliser des sources multiples d'information biométrique dans le but d'établir l'identité d'un individu avec une précision très élevée. Les systèmes biométriques multimodaux combinent l'évidence biométrique à partir de: plusieurs capteurs biométriques (ex., capteurs de visages 2D et 3D), plusieurs algorithmes (ex., correspondance d'empreintes digitales basée sur les minuties et les rides), plusieurs échantillons (ex., profils frontale, gauche et droit du visage), plusieurs instances (ex., iris droite et gauche) ou plusieurs traits (ex., visage et iris) afin d'améliorer la précision des systèmes biométriques. La fusion d'informations peut être accomplie à différents niveaux dans le système biométrique y compris le niveau capteur, le niveau caractéristique ou le niveau décision (Pour plus de détails sur la biométrie multimodale, voir la référence [40]).

1.7 Considération

Comme nous l'avons signalé à l'introduction, on s'intéresse dans le cadre de cette thèse à deux modalités biométriques qui sont: le visage et l'oreille. *Pourquoi ce choix ?* Tout d'abord, le visage de l'être humain est considéré comme la modalité la plus naturelle pour identifier les personnes; ce modèle correspond exactement à la manière de reconnaissance utilisée par le système neurologique humain. Cette modalité est non-intrusive et très acceptable par le public. Son avantage le plus important réside dans le système d'acquisition; une simple caméra reliée à un ordinateur peut suffire, la capture se fait à distance et d'une manière secrète. Par exemple, l'identification des criminels et des suspects dans les aéroports et dans les autres secteurs publics par des vidéos surveillances, est un scénario typique dans lequel apparaît l'importance du visage. Cependant, jusqu'à présent il n'existe aucune méthode qui fournit une solution robuste à toutes les situations et pour les différentes applications. *Qu'est-ce qui rend l'identification par visage difficile ?* Le visage de l'être humain est une entité dynamique qui se varie considérablement à cause des changements en pose, expressions faciales, illumination, occlusion, complexité de l'arrière plan et d'autres facteurs tels que: l'âge, le changement de la coiffure et la présence des lunettes, moustaches et barbes etc. [41] Aujourd'hui, les systèmes biométriques basés sur le visage sont, malheureusement, vulnérables aux attaques malveillantes. Une attaque de type "*Spoofing attack*" se produit

quand une personne essaye, délibérément, de trafiquer les données, en tant qu'une autre personne et de gagner des avantages illégitimes [42].

D'autre part, l'oreille humaine est une modalité très récente dans le domaine de la biométrie; en effet, il n'existe jusqu'à présent aucun logiciel commercialisé basé sur cette modalité. Elle est considérée comme l'une des caractéristiques anatomiques humaines les plus stables. Les études embryologiques ont prouvé que les changements principaux de la forme de l'oreille se produisent avant l'âge de huit (08) ans et ensuite après l'âge de soixante-dix (70) ans [28]. L'oreille ne change pas considérablement durant la vie humaine, tandis que le visage se change avec l'âge d'une manière très significative par rapport aux autres parties du corps humain. Les caractéristiques du visage peuvent être changées en fonction des cosmétiques utilisés, la coiffure et la coupe des cheveux. En plus, les visages humains se changent aussi avec les émotions et les différentes expressions de l'esprit comme la tristesse, la joie, la crainte, ou la surprise. En revanche, les caractéristiques de l'oreille sont riches, fixes et inchangeables par les émotions. Dans le processus d'acquisition et contrairement aux systèmes d'identification par visages, les images d'oreilles ne peuvent pas être masquées par les lunettes, les barbes ou les moustaches. Cependant, une occlusion partielle par cheveux ou par boucles est possible. Il est très important de noter que la biométrie par oreille est fortement acceptée par le public dans les applications de contrôle d'accès et dans la sécurité gouvernementale telles que: les programmes de visas et passeports. En outre, il n'y a aucun besoin de toucher le capteur et donc le problème d'hygiène ne se pose pas; l'acquisition des images peut être effectuée discrètement à distance et ne nécessite pas une coopération utilisateur-capteur. De plus, les images d'oreilles sont plus sécurisées que les images de visages, parce qu'il est difficile d'associer visuellement une image d'oreille à une personne donnée (en effet, la plupart des utilisateurs ne peuvent pas identifier leur propre image d'oreille). Par conséquent, les bases de données contenant les images d'oreilles ne doivent pas être plus sécurisées que les bases de données contenant les images de visages, puisque le risque d'attaque dans le premier cas est moins inférieur. Les images d'oreilles peuvent être acquises simultanément avec les images de visages et employées ensemble afin d'améliorer d'une manière significative la précision de la reconnaissance. Il est donc avantageux d'employer, dans un futur travail, le visage et l'oreille comme une pièce complémentaire d'information (multi-modalités), spécialement dans des applications de surveillance, suivi,

authentification personnelle continue et de lutter contre les attaques de type "*spoofing attacks*".

1.8 Biométrie cachée

Dans cette section, nous discutons une idée qui consiste à l'utilisation de la biométrie médicale pour des fins sécuritaires. En d'autres termes, au lieu de reconnaître un individu en utilisant son visage, empreinte digitale, iris, oreille, ou autres traits biologiques *visibles* et *accessibles*, *pourquoi ne pas utiliser quelques traits inaccessibles ? Pourquoi ne pas utiliser comme exemple: les motifs du cerveau ou la texture de l'os dans la sécurité biométrique ?* Ce nouveau type de biométrie, qui est en cours d'exploration, s'appelle "*la biométrie cachée*", elle consiste à utiliser des caractéristiques physiques et biologiques du corps humain, non visibles, sans traces et infalsifiables [43]. Comme il est facile de reconnaître un individu en utilisant des caractéristiques visibles et accessibles; ceci peut être simple, rapide, acceptable, et non coûteux. En plus, il est possible, dans certains cas, d'effectuer une identification à distance (i.e., reconnaissance en vidéo surveillance). Cependant, le défi majeur qui se pose en biométrie classique est que ces modalités sont sensibles à quelques attaques malveillantes (ex., *spoofing attacks* (fausse empreinte, fausse voix, faux visage)) et aussi sensibles aux blessures, coupures et contusions. Avec l'utilisation de la biométrie cachée, aucune personne ne peut modifier, délibérément, la géométrie et la texture de son cerveau ou la forme de certaines parties de son os.

Il semble clair, actuellement en 2015, que ce type de solutions reste inconvenable. Il est difficile d'employer ce type de biométrie pour sécuriser ou accéder à des ressources informatiques. La contrainte principale réside dans le système d'acquisition; la visualisation de la forme du cerveau nécessite l'utilisation des scanners MRI (Magnetic Resonance Imaging) et la visualisation du squelette du corps, contenant le crâne et les autres os, nécessite l'usage des scanners de rayons X (X-ray images). *Dans un futur proche, nous espérons que certains systèmes développés pour les applications médicales soient aussi utilisés dans la sécurité biométrique pour accéder à quelques ressources de haute sécurité.* Comme exemple, dans le but d'augmenter la sécurité dans quelques aéroports, les scanners de rayons X-ray à faible résolution peuvent être utilisés pour contrôler les frontières (initialement sont employés pour détecter les objets interdits dans les bagages). Nous espérons qu'un jour ce concept soit réalisé et adopté dans des applications spécifiques consacrées à la biométrie, par le

développant des scanners appropriés (en rapidité, acceptabilité et santé) afin d'appliquer efficacement et précisément *la biométrie cachée* en utilisant quelques parties invisibles du corps humain. Nous présentons par la suite quelques idées et travaux réalisés dans ce domaine.

1.8.1 Utilisation des images MRI en biométrie

L'objectif de cette section est de présenter l'utilisation des images MRI du cerveau humain dans l'identification des individus. Les figures 1.6.a et 1.6.b montrent deux images de deux cerveaux différents. Il est facile de les différencier visuellement. Ainsi, l'analyse de leurs textures par l'utilisation de quelques algorithmes de traitement d'images, peut être efficacement implémentée dans un processus de reconnaissance automatique. Dans ce contexte, la plupart des slices (i.e., coupures, voir figure 1.7.a) de l'image volumétrique MRI, à un niveau approprié, peuvent être considérées comme une signature; plusieurs approches peuvent être employées pour extraire des caractéristiques saillantes. Il a été démontré, empiriquement, que l'utilisation d'un seul niveau de coupure à un niveau donné est suffisante pour distinguer les individus [44]. A partir de la surface d'un objet, nous pouvons aussi extraire quelques paramètres capables de caractériser le cerveau humain.

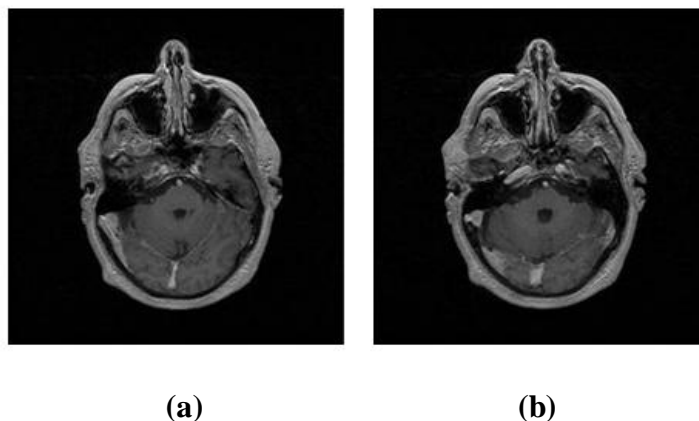


Figure 1. 6 : Deux images MRI du cerveau humain qui montrent une différence visuelle entre deux individus : (a) individu 01, (b) individu 02.

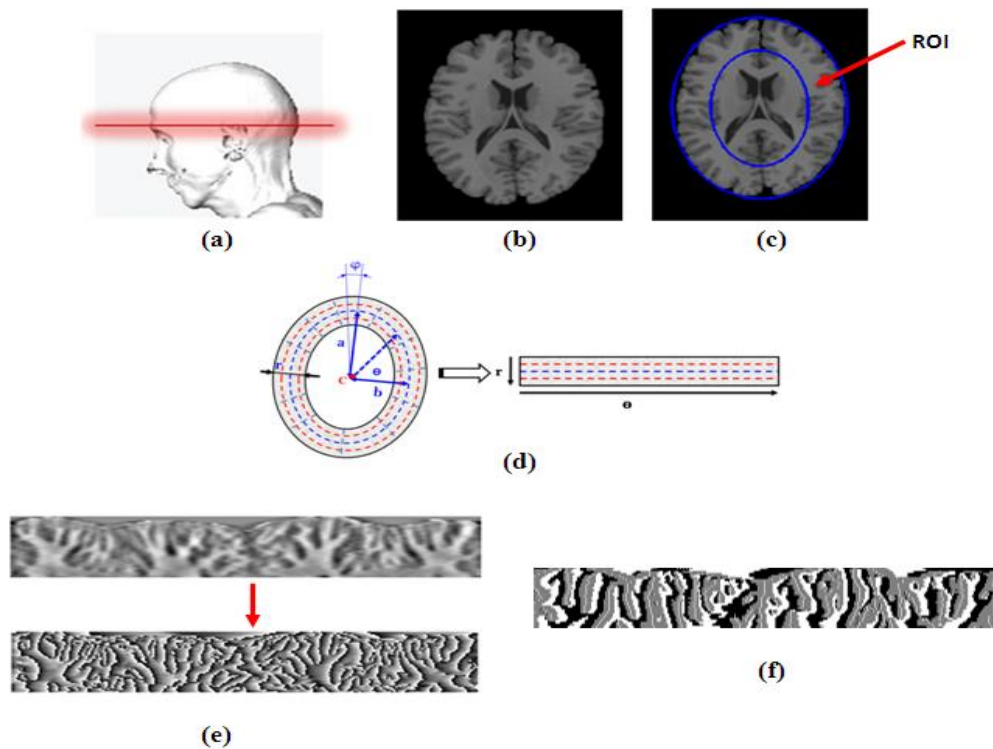


Figure 1.7 : Biométrie cachée par images 2D du cerveau humain. (a) Sélection du slice dans un niveau approprié du vertex. (b) Image MRI correspondante. (c) Extraction de la région ROI en utilisant une approche similaire à celle utilisé en identification par iris. (d) Dépilement de la bande elliptique (e) Normalisation de la bande elliptique. (f) Le code bar du cerveau (modèle) appelé *le code cerveau*.

Aloui et al. (2011) [45] ont développé un système de vérification d'identité basé sur la texture du cerveau; la méthode proposée est très similaire à celle développée par *Daugman* [19] (utilisée en reconnaissance par iris). Ils ont extrait une slice à partir de l'image MRI volumétrique, localisée à un niveau approprié du vertex comme présenté dans la figure 1.7.a. La région ROI (*Region of Interest*, voir figure 1.7.c) a été considérée comme une bande elliptique; cette région est délimitée par deux ellipses, une ellipse interne et une ellipse externe. Spécifiquement, cette bande contient le gyrus et la sulcature qui sont très variables d'un individu à un autre. Ensuite, la bande elliptique extraite est transformée en une image bidimensionnelle afin de faciliter les traitements, cette transformation peut être vue comme étant une tentative d'étendre ou de déplier la bande elliptique comme étant une feuille de caoutchouc (voir figure 1.7.d). Finalement et après un prétraitement, le processus de codage produit un code binaire contenant l'information nécessaire pour modéliser le cerveau qui a été nommé: *Le code Cerveau*.

Le concept de la biométrie cachée par l'utilisation des images MRI peut être généralisé en traitant tout le volume du cerveau. Considérant la représentation 3D comme présentée dans la figure 1.8, de nombreux paramètres peuvent être analysés. Si nous considérons les paramètres géométriques, la forme 3D du cerveau peut être utilisée et si nous considérons l'analyse de texture, nous pouvons utiliser la structure de convolution du cerveau [46]. Cette technique est jusqu'à présent en cours de recherche.



Figure 1. 8 : Biométrie cachée en 3D. Toute la forme du cerveau ou la convolution du cerveau peut être modélisée.

1.8.2 Utilisation des images X-ray en biométrie

La biométrie cachée a été aussi prolongée vers l'utilisation des images de rayons X (X-ray images), particulièrement avec des images de rayons à faible résolution. Comme illustration, nous présentons, dans les figures 1.9 (a-b), deux images X-ray de poumons qui correspondent à deux individus différents. Les différences en termes de textures et en morphologies peuvent être remarquées facilement. Dans ce cas, l'extraction des caractéristiques par l'utilisation de quelques techniques appropriées de traitement d'images peut être facilement employée pour différencier entre les individus [47]. Une autre application potentielle de la biométrie cachée qui utilise les images X-ray de la main est présentée dans les figures 1.9 (c-d). Cette méthode est plus efficace et robuste par rapport à la biométrie classique qui utilise les empreintes palmaires. Dans ce type d'application, les paramètres géométriques peuvent être facilement extraits et modélisés. Pour chaque individu, un code d'identification peut être extrait en mode d'apprentissage, et ensuite, ce code va être utilisé en mode de reconnaissance. Le nombre de paramètres extraits dépend de chaque approche prise en considération. Par exemple, à partir de l'image 1.9.d, nous pouvons extraire plusieurs paramètres géométriques qui correspondent à chaque phalange (ex., distal, moyen et proximal) dénotée par $\{a_1, a_2, a_3, a_4, a_5, b_1, \dots, d_1,$

d_2, \dots }. Ce vecteur peut être utilisé comme une signature intéressante pour reconnaître les individus [47].

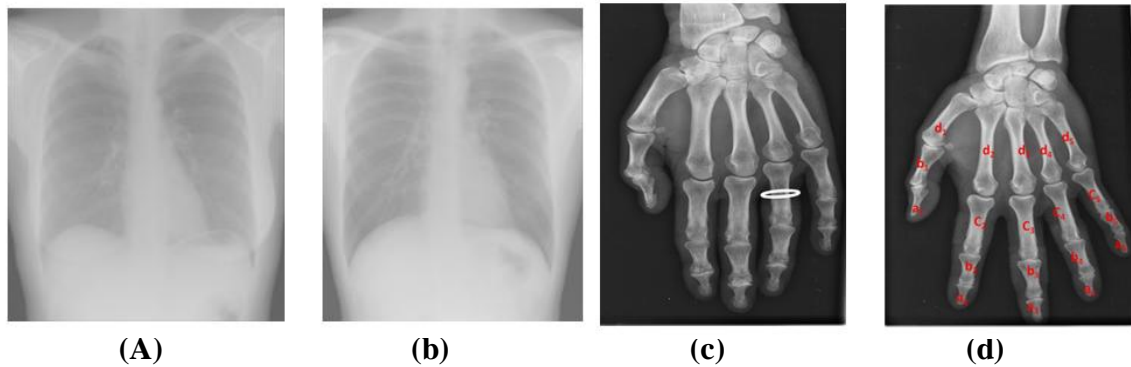


Figure 1. 9 : Biométrie cachée appliquée sur les images X-ray. (a) et (b) deux images X-ray des poumons de deux individus différents. (c) et (d) deux images X-ray des mains des deux individus différents.

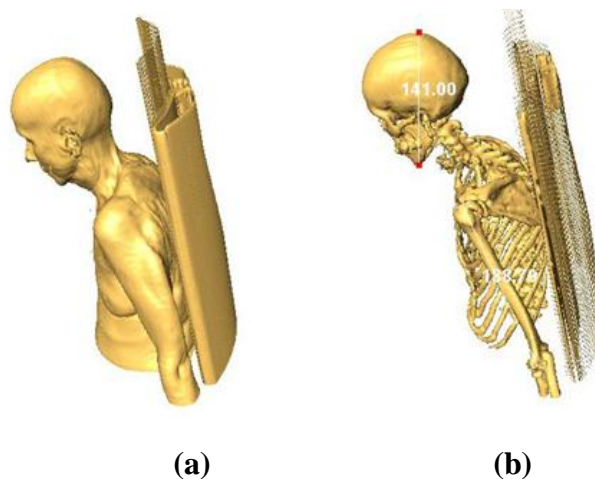


Figure 1. 10 : Biométrie cachée par les images X-ray appliqué au corps humain. (a) Image X-ray du corps humain. (b) Image X-ray du corps humain après seuillage.

Selon le principe de la biométrie cachée, par les images X-ray, le processus peut être généralisé si nous considérons l'analyse complète du corps humain. Comme présenté dans la figure 1.10, le squelette humain peut être extrait par une technique simple de seuillage (segmentation 3D) dont le contraste entre l'os et les autres textures est très haut. Le seuil est défini à partir de l'histogramme en utilisant des techniques de segmentation bien connues (ex., la logique floue, le seuillage adaptatif, etc.) et par conséquent, quelques détails valables à utiliser dans le contexte de la biométrie, peuvent être efficacement exploités à partir du squelette ou de quelques os spécifiques, par l'analyse de leur morphologie. En particulier,

plusieurs mesures, caractéristiques et paramètres peuvent être obtenus si nous considérons les vues axiales, coronales et sagittales du crâne comme présenté dans la figure 1.10.b [48,49].

1.9 Conclusion

A travers ce premier chapitre, nous avons présenté un état de l'art sur la biométrie, ses propriétés, le principe de fonctionnement des systèmes biométriques, les différentes modalités ainsi que les critères d'évaluation des performances de ce type de systèmes. Ensuite, nous avons mis en évidence une comparaison entre ces modalités biométriques, tout en accordant une attention particulière à la reconnaissance par *visage* et *oreille*, puisqu'elles constituent un bon choix, en termes de praticabilité, robustesse, acceptabilité et nouveauté, afin d'évaluer les descripteurs de texture locaux proposés dans cette thèse. Finalement, nous avons terminé le chapitre par une brève présentation de la *biométrie cachée* comme nouvel axe de recherche en criminalistique et en sécurité biométrique, qui constitue un défi très important que nous voulons exploiter dans un futur travail, par l'application et le développement des descripteurs de texture locaux proposés. Dans les deux chapitres suivants, nous évoquerons un état de l'art en reconnaissance de visages / oreilles, respectivement.

Chapitre 2

Reconnaissance Biométrique par Visages

2.1 Introduction

Comme l'une des applications les plus réussites dans le domaine de l'analyse et la compréhension de l'image; la reconnaissance du visage a reçu beaucoup d'attention, particulièrement dans les deux dernières décennies. Au moins deux raisons expliquent cette tendance: la première est la large gamme d'applications commerciales et légales, et la deuxième réside dans la disponibilité de technologies applicables. Même si les systèmes courants de reconnaissance par machines ont atteint un certain niveau de maturité, leur succès est limité aux conditions imposées par plusieurs applications réelles. Par exemple, la reconnaissance des images faciales acquises dans un environnement extérieur avec des changements d'éclairage et/ou pose, reste un problème largement non-résolu. En d'autres termes, les systèmes actuels sont encore loin des capacités du système de la perception humaine.

Dans ce chapitre, nous présentons un aperçu critique sur la recherche statique et mobile en reconnaissance de visages. Nous sommes motivés, dans l'écriture de ce chapitre, par deux points fondamentaux, qui sont: (i) fournir un examen mis-à-jour sur la littérature existante et (ii) offrir quelques perspicacités sur les études de reconnaissance de visages par machines. Pour ce faire, nous classons non seulement les techniques de reconnaissance existantes mais nous présentons aussi des descriptions détaillées sur les méthodes représentatives au sein de chaque catégorie. En plus, des sujets relevant sont couverts, tels que: les études neuropsychologiques, l'évaluation du système et des issues sur l'éclairage et les variations de la pose.

2.2 Le visage: le moyen naturel pour la reconnaissance humaine

Depuis l'avènement de la photographie, les agences gouvernementales et les organismes privés ont gardé des collections de photos faciales pour les personnes (ex., les documents d'identifications personnelles, les passeports, les cartes d'adhésion, etc.). Avec la large

utilisation des caméras numériques et les téléphones intelligents, des images de visages peuvent être facilement générées chaque jour. En outre, aujourd'hui ces images peuvent être transmises et partagées rapidement par les réseaux sociaux fortement développés tels que: *Facebook* et *Twitter*. Donc, le visage est presque le trait biométrique le plus commun et le plus familier dans nos vies quotidiennes. Il y a plus de possibilités afin d'acquérir et analyser les images de visages d'une personne interrogée (ex., suspect, témoin ou victime) à des fins d'enquêtes judiciaires.

La reconnaissance de visages a une longue histoire; elle suscite les intérêts de recherches des neurologistes, psychologues et informaticiens [50]. En comparaison à d'autres traits biométriques, le visage n'est pas *parfait*. Par exemple, il est généralement moins précis que d'autres formes de données biométriques telles que l'empreinte digitale et peut être potentiellement affecté par des produits cosmétiques. Cependant, le visage a ses propres avantages qui lui font l'un des traits biométriques les plus privilégiés pour la reconnaissance humaine:

- a. Nature biologique:** le visage est une caractéristique biométrique très pratique, utilisée par les humains dans la reconnaissance des personnes, ce qui en fait probablement le trait biométrique le plus courant pour des fins d'authentification et d'autorisation. Par exemple, dans le contrôle d'accès, il est facile pour les administrateurs de suivre et analyser les personnes autorisées, en utilisant ses données de visages après authentification. L'aide des utilisateurs ordinaires (ex., les administrateurs dans ce cas) peut améliorer la fiabilité et l'applicabilité des systèmes de reconnaissance. Alors que la reconnaissance par empreintes digitales ou iris exige un expert ayant des compétences professionnelles pour fournir une confirmation fiable.
- b. Non-intrusion:** contrairement aux collections d'empreintes digitales ou iris, les images faciales peuvent être facilement acquises à distance et sans contact physique. Les gens se sentent plus confortables pour l'utilisation du visage comme un identifiant dans la vie quotidienne. Un système de reconnaissance par visages peut collecter des données biométriques d'une manière amicale facilement acceptée par le public.
- c. Moins de coopération:** en comparaison à l'iris ou l'empreinte digitale, la reconnaissance du visage exige moins de coopération par l'utilisateur. Dans certaines

applications particulières telles que la surveillance, un système de reconnaissance de visages peut identifier une personne sans aucune participation active de l'objet.

Les premières tentatives que nous connaissons pour identifier un objet consistent à comparer une partie de la photographie faciale, elles ont été rapportées la première fois dans un tribunal *Britannique* en 1871 [51]. La reconnaissance du visage est l'une des tâches les plus importantes dans les enquêtes légales s'il y a n'importe quel matériel vidéo ou image fournie sur une scène de crime. Les experts légaux effectuent un examen manuel sur les images faciales afin de faire une correspondance avec celles d'un suspect. L'utilisation des systèmes de reconnaissance de visages automatisés, améliore non seulement l'efficacité du travail judiciaire mais normalise également le processus de la comparaison.

2.3 Inspirations neuropsychologiques en reconnaissance de visages

Le processus de la reconnaissance humaine utilise un large spectre de stimulus obtenu à partir de plusieurs, si pas tous, les sens (visuel, auditif, olfactif, tactile, etc.). Dans de nombreuses situations, la connaissance contextuelle est également appliquée, par exemple, les environnements jouent un rôle important en reconnaissance de visages par rapport à l'endroit sensé. Il n'est pas utile d'essayer même de développer un système utilisant une technologie existante, qui imite les capacités remarquables de l'être humain en reconnaissance de visages. Cependant, le cerveau humain a ses limitations, surtout dans le nombre total des personnes qu'il peut exactement « *rappeler* ». L'un des principaux avantages d'un système informatique réside dans sa capacité à gérer un grand nombre d'images faciales. Dans la plupart d'applications, les images sont disponibles seulement sous la forme d'une unique ou de multiples points des vues, de données d'intensités 2D, de sorte qu'uniquement les entrées aux algorithmes de reconnaissance automatique par visages soient visuelles. Pour cette raison, la littérature examinée dans cette section est limitée à des études liées à la perception visuelle des visages par l'être humain.

De nombreuses études en psychologie et en neurosciences ont une pertinence directe avec les ingénieurs intéressés par la conception des algorithmes ou des systèmes de reconnaissance automatique de visages. Par exemple, les résultats de *Bruce* [52] (1988) et *Shepherd et al.* [53] (1981) en psychologie, concernant l'importance relative de différentes caractéristiques

faciales sont notées dans la littérature de l'ingénierie par *Etemad et Chellappa* [54] (1997). D'autre part, les systèmes machines fournissent des outils pour entreprendre des études en psychologie et en neuroscience [55,56]. Par exemple, une explication possible de l'ingénierie sur les effets de l'éclairage au fond a été étudiée par *Johnston et al.* [57] (1992), ils ont suggéré que: quand la direction réelle de l'éclairage est opposée à la direction habituellement supposée, le système récupère des informations structurelles incorrectes et par conséquent il rend la reconnaissance de visages plus dure.

Considération: Un examen général sur les études détaillées en psychologie et en neuroscience dépasse la portée de notre étude. Ici, nous récapitulons seulement les résultats qui sont potentiellement appropriés à la conception du système de reconnaissance de visages. Les questions qui peuvent être intéressantes durant la phase de l'implémentation sont:

- **Est-ce que la reconnaissance du visage est un processus dédié ?** *Biederman et Kaloscai* [58] (1998), *Ellis* [59] (1986), *Gauthier et al.* [60] (1999) et *Gauthier et Logothetis* [61] (2000) ont suggéré: "Nous pensons traditionnellement que la reconnaissance du visage est un processus dédié qui est différent aux autres tâches de reconnaissance d'objets". L'évidence pour l'existence d'un système de traitement consacré au visage provient de plusieurs sources [59]. (i) Les visages sont plus faciles à retenir par les humains que les autres objets lorsqu'ils sont présentés dans une orientation verticale. (ii) Les patients de prosopagnosie ne peuvent pas reconnaître des visages auparavant familiers. Ils reconnaissent les personnes par leurs voix, couleur des cheveux,...etc. Il faut noter que les patients de prosopagnosie peuvent reconnaître si un objet donné est un visage ou non, mais ils ont des difficultés à identifier les visages. Sept (07) différences entre la reconnaissance du visage et la reconnaissance d'objet peuvent être récapitulées [58] en se basant sur des épreuves empiriques: (1) les effets de la configuration (liés au choix de différents types de systèmes de reconnaissance par machine), (2) l'expertise, (3) les différences verbalisables, (4) la sensibilité au contraste de polarité et la direction de l'illumination (liées aux problèmes d'illumination dans les systèmes de reconnaissance par machine), (5) la variation métrique, (6) la rotation en profondeur (liée aux problèmes de la variation de pose dans les systèmes de reconnaissance par machine) et (7) la rotation dans le plan ou l'inversement du visage. Contrairement à ce que nous croyons traditionnellement, certaines découvertes récentes dans la neuropsychologie humaine et la neuro-imagerie suggèrent que la reconnaissance du

visage ne peut pas être unique. Selon *Gauthier et Logothetis* [61] (2000), des études récentes en neuro-imagerie chez les humains indiquent que le niveau de la catégorisation et de l'expertise interagit par activation afin de produire la spécification de visages dans le gyrus fusiforme moyen (Le gyrus fusiforme ou gyrus occipito-temporel se situe sur la surface ventro-médiane des lobes temporaux et occipitaux, voir figure 2.1, ceci est considéré comme un axe de recherche en neuroscience). Par conséquent, il est possible que le système de codage utilisé pour les visages puisse être également utilisé pour d'autres classes ayant des propriétés similaires (Ex., la reconnaissance de visages familiers contre visages non-familiers).

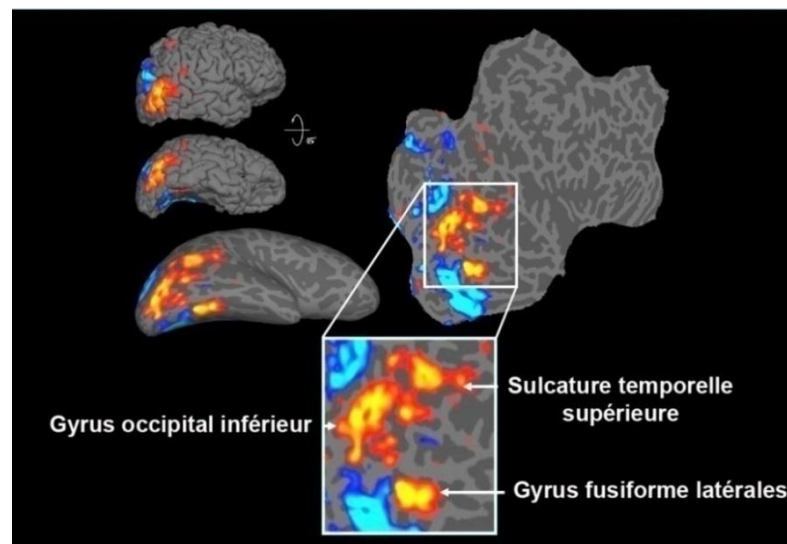


Figure 2. 1 : Les régions d'activation du cerveau répondants principalement au stimulus de visage (motifs rouge au jaune) ou à des photos de maisons (motifs bleu) (Reproduite de [62]).

- **Est-ce que la perception du visage est le résultat d'une analyse holistique ou basée sur les caractéristiques ?** *Bruce* [52] (1988) et *Bruce et al.* [63] (1998) ont suggéré: Les deux informations: globale (holistique) et locale (basée sur les caractéristiques) sont cruciales pour la perception et la reconnaissance du visage. Les études indiquent que la description locale présente plus de détails précis. Si les caractéristiques dominantes sont bien présentées, la description holistique ne peut pas être utilisée. Par exemple, dans les études de rappel de visages, les humains se concentrent rapidement sur les caractéristiques bizarres telles que: grandes oreilles, nez tordu, yeux écarquillés,...etc. Une des preuves les plus fortes pour soutenir l'idée qui suggère que la reconnaissance du visage implique un traitement plus complexe et holistique par rapport à la reconnaissance des autres objets, réside dans l'effet

de l'inversion du visage, dans lequel un visage inversé est beaucoup plus difficile à reconnaître qu'un visage normal (démonstré la première fois par Yin [64] (1969)). Un excellent exemple est donné par Barlett et Searcy [65] (1993) par l'utilisation de «l'illusion de Thatcher [66]»; dans cette illusion, les yeux et la bouche d'un visage exprimant, sont excisés et inversés, le résultat semble grotesque dans une vue verticale; cependant, en présence de l'inversion, la perception du visage reste normale en apparence, tandis que l'inversion des caractéristiques internes n'est pas facilement remarquée, voir figure 2.2.



Figure 2. 2 : L'illusion de *Thatcher* montre que les distorsions flagrantes sur les informations de base du visage sont rarement remarquées lorsque le visage est inversé (Reproduite de [64]).

- **Classement des caractéristiques faciales par ordre d'importance:** Bruce [52] (1988) et Shepherd et al. [53] (1981) ont suggéré: Les cheveux, le contour du visage, les yeux, et la bouche (pas nécessairement dans cet ordre) sont considérés comme des composantes très importantes afin de percevoir et rappeler les visages [53]. Plusieurs études ont prouvé que le nez joue un rôle insignifiant; cela peut être dû au fait que toutes les études sont réalisées avec des images frontales. Il est à noter que les profils peuvent être importants dans les applications de type "*Mugshot-ID*", où ils peuvent être extraits à partir de la vue de côté, comme montré dans la figure 2.3, une forme distinctive du nez peut être plus importante que les yeux ou la bouche [52]. Un autre résultat de certaines études montre que les caractéristiques externes et internes sont importantes pour la reconnaissance de visages célèbres (ce résultat ne fonctionne pas avec les visages non-familiers); les caractéristiques internes sont les plus dominantes dans la reconnaissance de visages familiers. Shepherd et al. [53] (1981) ont également constaté que la partie supérieure du visage est plus utile que la partie inférieure pour la reconnaissance. Le rôle des attributs esthétiques tels que la beauté, l'attractivité et/ou l'agrément a été aussi étudié, avec la conclusion que: plus les visages sont

attractifs, plus leurs taux d'identification sont meilleurs; les visages moins attractifs viennent ensuite, suivis des visages de mi-portés, en terme de facilité à reconnaître.



Figure 2. 3 : Exemple d'une application de type "Mugshot-ID".

- **Caricatures:** *Brennan* [67] (1985), *Bruce* [52] (1988) et *Perkins* [68] (1975) ont suggéré: Une caricature peut être formellement définie [68] comme: « un symbole qui exagère les mesures relatives à n'importe quelle mesure qui varie d'une personne à une autre ». Par exemple, la longueur du nez est une mesure variante d'une personne à une autre et peut être utile comme un symbole pour caricaturer quelqu'un. Un algorithme standard de caricatures [67] peut être appliqué à différents types d'images (dessins de traits et photographies). Les caricatures aux dessins de traits ne contiennent pas autant d'informations que les photographies, mais elles parviennent à capturer les caractéristiques importantes d'un visage; les expériences basées sur des visages non ordinaires comparant l'utilité des caricatures aux dessins de traits et les dessins de traits non exagérés, favorisent nettement le premier modèle [52].
- **Visages distinctifs:** *Bruce et al.* [69] (1994) ont suggéré: Les études montrent que les visages distinctifs sont mieux conservés en mémoire, ils sont mieux et plus vite reconnus que les visages typiques. Cependant, si une décision doit être prise afin de savoir qu'un objet est un visage ou non, celle-ci prend plus de temps pour reconnaître un visage atypique qu'un visage typique. Cela peut être expliqué par les différents mécanismes utilisés pour la détection et la reconnaissance.
- **Quel est le rôle de l'analyse fréquentielle en reconnaissance du visage ?** *Ginsburg* [70] (1978), *Harmon* [71] (1973) et *Sergent* [72] (1986) ont suggéré: Les études antérieures [70,71] ont conclu que l'information dans les basses bandes de fréquences joue un rôle

dominant dans la reconnaissance du visage. Les études récentes [72] montrent qu'en fonction de la tâche spécifique de la reconnaissance, les composantes à basses et hautes bandes de fréquences peuvent jouer des rôles différents. Par exemple, la classification du genre peut être accomplie avec succès en utilisant uniquement les composantes à basses fréquences. Tandis que, la reconnaissance de la personne nécessite en plus, l'utilisation des composantes à hautes fréquences [72]. Les composantes à basses fréquences contribuent à la description locale, alors que les composantes à hautes fréquences contribuent aux détails additionnels pour la reconnaissance.

- **Est-ce que la reconnaissance est liée à un point-de-vue unique ?** *Biederman* [73] (1987), *Hill et al.* [74] (1997) et *Tarr et Bulthoff* [75] (1995) ont suggéré: Beaucoup de travaux sur la reconnaissance visuelle d'objet (ex., *Biederman* [73] (1987)) sont moulés dans un cadre théorique introduit par *Marr* [76] (1982), dans lequel différentes vues d'objets sont analysées d'une manière permettant d'englober, en grande partie, les caractéristiques de cet objet à un point-de-vue invariant. Récemment, il y a eu un certain débat concernant la connaissance si la reconnaissance d'un objet, au niveau du cerveau, est effectuée à un point-de-vue invariant ou non [75]. Certaines expériences suggèrent que la mémorisation du visage est très dépendante à un point-de-vue unique [74].
- **Effet du changement de l'éclairage:** *Bruce et al.* [63] (1998), *Hill et Bruce* [77] (1996) et *Johnston et al.* [57] (1992) ont suggéré: Depuis longtemps, il a été observé d'une manière informelle qu'il est difficile de reconnaître les négatifs photographiques des visages. Cependant, peu de travaux relatifs ont exploré pourquoi il est difficile de reconnaître les images négatives des visages ? Les expériences de *Johnston et al.* [57] (1992) ont été menées dans le but d'explorer si les difficultés liées aux images négatives et inversées des visages surgissent, parce que chacune de ces manipulations inverse la direction apparente de l'éclairage et donne l'impression que l'image d'un visage éclairé du haut est éclairée par le bas. Il a été démontré par *Johnson et al.* [57] (1992) que l'éclairage inférieur (au fond) rend, en effet, l'identification de visages familiers plus difficile. *Hill et Bruce* [77] (1996) ont présenté que l'importance de l'éclairage supérieur pour la reconnaissance du visage est démontrée par l'utilisation d'une tâche différente (correspondance par des images de visages extérieures pour déterminer si elles sont identiques).
- **Reconnaissance de visages en mouvement:** *O'Toole et al.* [78] (2002), *Bruce et al.* [63] (1998) et *Knight et Johnston* [79] (1997) ont suggéré: Des études récentes [79] ont prouvé

qu'il est plus facile de reconnaître les visages célèbres lorsqu'ils sont apparus dans des séquences mobiles que dans des photographies statiques. Cette observation a été étendue afin de montrer que le mouvement contribue dans l'identification de visages familiers, présentés sous une gamme de différents types de dégradations: nié, inversé ou seuillé [63]. Encore plus intéressant, cette observation semble être avantageuse même si le contenu de l'information mobile est assimilé dans les conditions de la comparaison statique. Cependant, les expériences avec des visages non-familiers suggèrent qu'il n'y a aucun avantage additionnel avec le visionnement animé par rapport aux images statiques.

- **Expressions faciales:** *Bruce* [52] (1988) a suggéré: En se basant sur des études neuropsychologiques, il semble clair que l'analyse des expressions faciales est accomplie en parallèle avec la reconnaissance du visage. Quelques patients de prosopagnosie, qui ont des difficultés à identifier des visages familiers, néanmoins peuvent reconnaître les expressions en se basant sur les émotions. Les patients qui souffrent du « syndrome cérébral organique », souffrent d'une mauvaise analyse de l'expression, mais exécutent bien la reconnaissance du visage. De même, la séparation entre la reconnaissance du visage et les tâches du traitement visuel focalisé (ex., la recherche de quelqu'un avec une moustache épaisse) est demandée.

2.4 Reconnaissance de visages dans le domaine de l'ingénierie

Comme illustré dans le schéma de la figure 2.4, le problème de la reconnaissance automatique de visages implique trois étapes principales: (i) détection et normalisation approximative des visages, (ii) extraction des caractéristiques et normalisation précise des visages et (iii) classification (vérification / identification). Parfois, quelques étapes ne sont pas totalement séparées; par exemple, les caractéristiques faciales (yeux, nez et bouche) utilisées dans la reconnaissance sont souvent employées durant la détection de visages. La détection et l'extraction des caractéristiques peuvent être réalisées simultanément, comme indiqué dans le schéma de la figure 2.4. Selon la nature de l'application, par exemple, la taille de la base de données, le désordre et la variabilité de l'arrière plan, le bruit, l'occlusion, la vitesse et certaines tâches secondaires peuvent être considérées comme obstacles très provocants.

Bien que les systèmes entièrement automatiques de reconnaissance de visages doivent exécuter les trois étapes citées précédemment, la recherche dans chaque étape est considérée comme un sujet critique; non seulement par ce que les techniques utilisées pour chaque étape individuelle doivent être améliorées, mais également parce qu'elles sont critiques dans

beaucoup d'applications différentes (Voir figure 2.4). Par exemple, la détection du visage est essentielle pour initialiser le suivi du visage et l'extraction des caractéristiques faciales est nécessaire pour reconnaître l'état émotionnel de la personne, qui est à son tour essentiel dans les systèmes d'interaction homme-machine (IHM). L'isolement de chaque étape facilite l'évaluation et l'évolution en état de l'art des techniques composantes. Les premières techniques de détection de visages peuvent seulement manipuler des visages simples ou frontaux bien-séparés dans des images contenant un arrière plan simple. Aujourd'hui, les algorithmes de détection peuvent localiser les visages et estimer leurs poses avec des arrière plans encombrés [80–82]. Des recherches extensives sur chaque étape sont effectuées et des aperçus relevant pour chaque étape sont apparus.

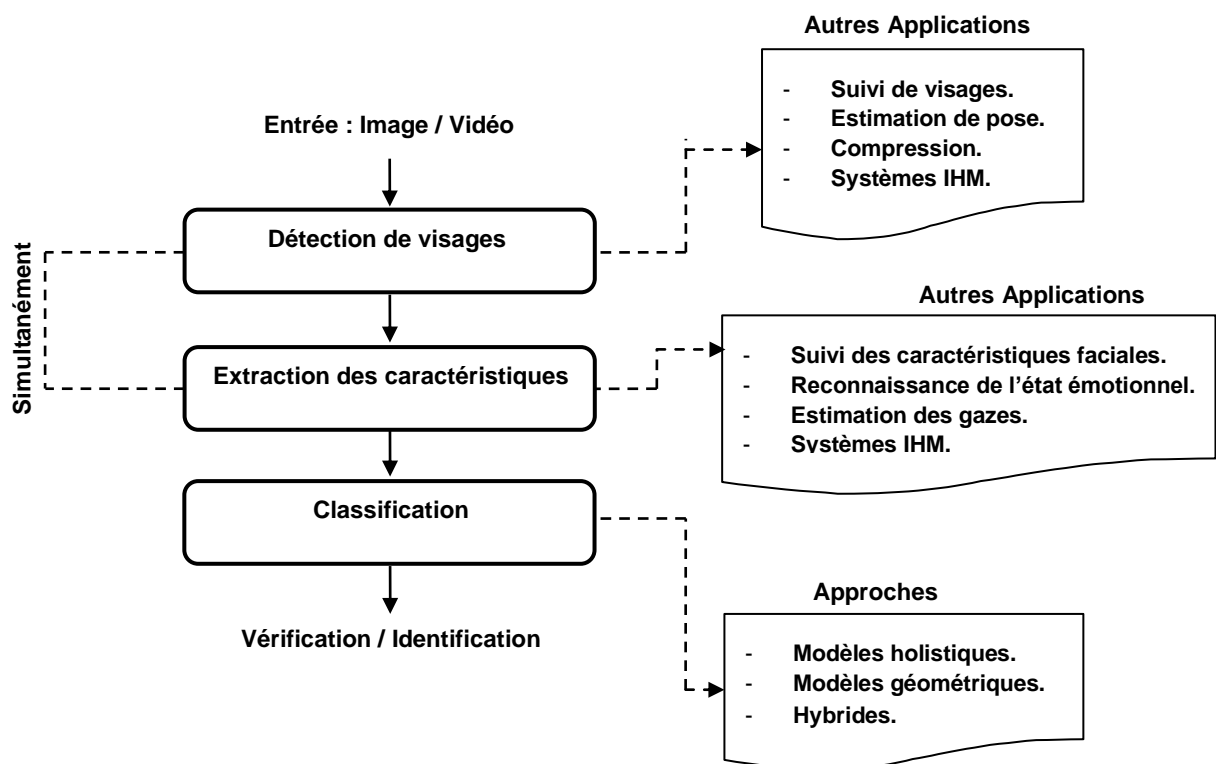


Figure 2. 4 : Configuration générique d'un système de reconnaissance automatique de visages.

Dans cette section, nous examinons l'état de l'art en reconnaissance de visages dans la littérature de l'ingénierie. Pour la bonne perfection, nous présentons des sommaires sur la recherche en détection/segmentation et extraction des caractéristiques. Ensuite, nous donnons un examen détaillé sur les travaux récents en reconnaissance de visages.

2.4.1 Etapes clés pour la reconnaissance: Détection de visages et extraction des caractéristiques

La première étape dans n'importe quel système de reconnaissance automatique par visages est la détection de visages dans l'image. Ici, nous fournissons seulement un sommaire sur ce sujet tout en accordant une attention particulière aux méthodes récentes. Après la détection d'un visage vient la tâche d'extraction des caractéristiques, celle-ci consiste à obtenir des caractéristiques saillantes qui sont introduites par la suite dans un système de classification. Selon le type du système de classification, les caractéristiques peuvent être *locales* telles que les lignes ou les points fiduciels, ou *faciales* telles que les yeux, le nez et la bouche. La détection peut également utiliser ces caractéristiques, dans ce cas, ces dernières sont utilisées simultanément dans l'étape de détection et l'extraction. Ainsi, l'extraction des caractéristiques est une étape clé dans les systèmes d'animation et d'identification des expressions faciales.

Sans tenir compte de la localisation des caractéristiques, la détection de visages est considérée comme réussie, si la présence et la localisation approximative d'un visage sont correctement identifiées. Alors, sans détection précise et sans localisation des caractéristiques, une dégradation notable dans la performance de la reconnaissance sera observée [83]. La relation étroite entre l'extraction des caractéristiques et la reconnaissance du visage nous motive à donner un aperçu sur les méthodes d'extraction des caractéristiques communément utilisées dans ce domaine.

2.4.2 Segmentation / Détection de visages (Sommaire)

Au milieu des années 90, la plupart des travaux sur la détection sont focalisés sur la segmentation d'un visage unique à partir d'un arrière plan simple ou complexe. Parmi ces approches, nous trouvons des méthodes basées sur: la connaissance d'un modèle entier de visage, la mise en correspondance avec des modèles déformables, la couleur de la peau humaine et les réseaux de neurones.

Des progrès significatifs ont été faits au cours de ces dernières années pour réaliser la détection automatique de visages sous diverses conditions. La comparaison des méthodes basées sur les caractéristiques (ex., couleur de la peau) à celles qui sont basées sur la correspondance; l'apparence, c'est à dire les méthodes basées sur l'image [84,85] qui forment des systèmes machines à partir d'un grand ensemble d'échantillons, achève les meilleurs

résultats. Cela n'est pas surprenant puisque les objets de visages sont compliqués, très semblables entre eux et différents par rapport aux objets de non-visages. Grâce à une formation approfondie, les ordinateurs peuvent être tout à fait bons pour détecter les visages.

Plus récemment, la détection de visages sous rotation approfondie a été étudiée. Ces approches sont basées sur l'apprentissage à partir des échantillons de multiples vues [82,86]. En comparaison avec les méthodes basées sur les caractéristiques invariantes, les méthodes de détection et reconnaissance basées sur les multiples vues, semblent être capables d'achever des résultats meilleures quand l'angle de rotation hors plan est large (maximum 35°). Dans la communauté de psychologie, un débat similaire existe, autour de la question qui cherche à savoir si la reconnaissance du visage est dépendante à un point de vue invariant ou non. Les études dans les deux disciplines supportent l'idée qui stipule que pour les petits angles, la perception du visage, est une vue-indépendante tandis que pour les grands angles, c'est une vue-dépendante.

Dans un problème de détection, deux statistiques sont importantes: *le taux de vraies détections* rapportées dans la région de visages et *le taux de fausses détections* rapportées dans la région de non-visages. Un système idéal doit avoir un taux élevé de vraies détections et un taux faible de fausses détections. Dans la pratique, ces deux exigences sont contradictoires. Le fait de traiter le problème de détection de visages comme un problème de classification à deux classes; aide à réduire les fausses détections d'une façon dramatique tout en maintenant les vraies détections. Cela est achevé par des systèmes de recyclage avec des échantillons faussement positifs générés par des systèmes formés précédemment.

Contribution (Système proposé): Au cours de cette thèse, nous avons proposé un système hybride de détection de visages dans une image couleur et sous conditions variantes (illumination, pose, occlusion et taille du visage; non-contrôlées). Ce système est basé principalement sur une technique d'apprentissage automatique en utilisant la décision de trois réseaux de neurones, une méthode d'extraction et réduction des caractéristiques basée sur le principe de compactage d'énergie utilisant la transformée discrète en cosinus (DCT) et une technique de segmentation par la couleur de la peau humaine pour réduire l'espace de recherche dans l'image et pour accélérer le processus de détection. Un ensemble d'images (visages et non-visages) est utilisé pour entraîner les réseaux de neurones à séparer entre les deux classes. Les résultats expérimentaux ont montré que cette hybridation a apporté des

améliorations très significatives, en termes de taux de vraies détections, de la qualité de détection et de temps d'exécution [81,87].

2.4.3 Extraction des caractéristiques (Sommaire et méthodes)

Indépendamment des dispositifs d'acquisition exploités pour saisir les flux de l'image, une taxonomie simple qui peut être basée sur l'architecture informatique est appliquée afin d'extraire des caractéristiques distinctives, et probablement uniques pour l'identification ou la dérivation d'un modèle de description, qui sera utilisé ultérieurement en correspondances.

Les deux catégories algorithmiques principales, peuvent être définies sur la base de la relation entre l'objet et le modèle du visage, c'est à dire, l'algorithme est basé sur une représentation d'objet-centré (éco-centrique) ou sur une représentation de caméra-centrée (égocentrique). La première classe d'algorithmes s'appuie sur un modèle de visage complexe généralement 3D ou 2.5D; qui est fortement lié à la forme 3D du visage. Ces méthodes reposent sur une procédure plus complexe pour extraire les caractéristiques et construire le modèle du visage, mais elles ont l'avantage d'être intrinsèquement invariantes aux poses. Les algorithmes de type visage-centré les plus populaires sont ceux basés sur l'acquisition de données de visages 3D et les cartes de visages en profondeur.

La classe d'algorithmes égocentriques repose fortement sur l'information des structures au niveau de gris sur les images. Par conséquent, la représentation du visage est fortement variante à la pose, ainsi, le modèle est rigidement lié à l'apparition du visage plutôt qu'à la forme 3D du visage. Les algorithmes de type image-centrée les plus populaires sont les méthodes holistiques (sous-espaces), les méthodes basées sur les caractéristiques (géométriques) et les méthodes hybrides. A partir de ces classes d'algorithmes élémentaires, plusieurs élaborations ont été proposées. Parmi lesquelles, les méthodes du *noyau* ont grandement amélioré la puissance discriminative de plusieurs algorithmes égocentriques, au moment où les nouvelles techniques d'analyse de caractéristiques, telles que la représentation par le *motif binaire local* (LBP pour Local Binary Pattern) ont considérablement amélioré la vitesse et la robustesse des méthodes basées sur le filtre de *Gabor*. Les mêmes considérations sont valables pour les algorithmes éco-centriques, où les nouveaux descripteurs de la forme et les modèles 3D paramétriques, y compris la fusion des informations de la forme avec la texture 2D du visage, ont considérablement amélioré la précision des méthodes existantes.

a. Méthodes basées sur les sous-espaces (holistiques / globales)

Les algorithmes basés sur les sous espaces proviennent de la prétention qui suppose que n'importe quelle collection M d'images de visages, contient des redondances qui peuvent être éliminées par application d'une décomposition de tenseur. Cette procédure produit un ensemble de vecteurs basiques qui représente un espace de dimension inférieure tout en respectant l'ensemble original d'images. Etant donné l'ensemble des vecteurs basiques, chaque visage peut être reconstruit dans l'espace réduit. Pour faciliter le processus, chaque image de visage $N \times N$ est représentée comme un vecteur obtenu par l'alignement des rangés de l'image. La matrice résultante $(N \times N) \times M$ est décomposée pour obtenir les vecteurs basiques non-singuliers. La classification est habituellement effectuée par la projection d'une nouvelle image de visage acquise dans l'espace à faible dimension (sous-espace) et le calcul d'une mesure de distance avec toutes les classes représentées dans cet espace.

Plusieurs critères sont utilisés pour déterminer la base des espaces à faible dimension. Une classe d'algorithmes utilise les projections qui représentent mieux la population mais sans aucune information liée aux différentes classes. D'autres approches adressent plus explicitement la discrimination entre les classes. L'indépendance statistique dans l'espace des caractéristiques à faible dimension est également imposée pour trouver les projections linéaires. L'approche pionnière qui est l'approche de *l'analyse en composantes principales* (PCA: Principal Component Analysis ou *les visages propres* pour Eigen-faces) [88,89]; la représentation des visages propres, est basée sur une transformation linéaire qui maximise la dispersion de tous les exemples projetés par la dé-corrélation des données. Ceci correspondant à *une décomposition de valeurs singulières* (SVD: Singular Value Decomposition) appliquée sur l'ensemble d'images.

Etant donné un ensemble M d'images x_i , chaque image est composée de $N = n \times m$ pixels, l'espace basique de l'ensemble de données $M \times N$ est représenté par la matrice de composantes principales U . Ceci peut être calculé par application de la décomposition de valeurs singulières:

$$X = U.D.V \quad (2.1)$$

Où chaque rangée (ligne) de la matrice X est obtenue par une séquence de rangées groupées de chaque image, en plus le niveau de gris de chaque pixel est normalisé avec le

niveau moyen de l'image. La matrice D de $N \times N$, possède les valeurs singulières de X sur sa diagonale principale et les zéros ailleurs. Celles-ci correspondent aux valeurs propres de U . La principale limitation de l'approche PCA réside dans sa contrainte d'orthogonalité. L'ensemble calculé de vecteurs basiques est toujours orthonormé, par conséquent, ils sont incapables d'identifier aussi les directions de la variabilité maximale.

L'approche PCA est récemment prolongée vers un algorithme *non-linéaire* par l'utilisation d'une fonction de *noyau* (KPCA: Kernel Principal Component Analysis) [90,91].

La matrice de la factorisation non-négative (NMF: Non-Negative Matrix Factorization) [92] est une variation de la décomposition de PCA. La principale différence par rapport à la PCA est qu'elle n'autorise pas des éléments négatifs dans l'ensemble des vecteurs basiques et les coefficients de la combinaison linéaire. Une extension de la NMF qui donne des bases encore plus localisées en imposant des contraintes supplémentaires de la localité est: *l'analyse discriminante linéaire* (LDA: Linear Discriminant Analysis) [93,94]. *Bulhumeur et al.* [95] (1997) ont originalement proposé une méthode dérivée de la PCA et la LDA, où la réduction de la dimensionnalité avec PCA est suivie par un critère d'optimisation à l'aide de *la discrimination linéaire de Fisher* (FLD: Fisher's Linear Discriminant Optimization). FLD sélectionne le sous-espace Ψ qui maximise:

$$\frac{|\Psi^T S_b \Psi|}{|\Psi^T S_w \Psi|} \quad (2.2)$$

Où S_b est la matrice de dispersion inter-classes et S_w est la matrice de dispersion intra-classes.

$$S_b = \sum_{i=1}^m N_i (\bar{X}_i - \bar{X}) (\bar{X}_i - \bar{X})^T \quad (2.3)$$

$$S_w = \sum_{i=1}^m \sum_{x \in X_i} N_i (\bar{X}_i - \bar{X}) (\bar{X}_i - \bar{X})^T \quad (2.4)$$

Où S_b détermine la différence entre les images de différents objets, S_w définit la similarité entre les images d'un même objet. Afin d'éviter des singularités dans le calcul de S_b et S_w , une décomposition de PCA est d'abord appliquée sur l'ensemble de données X . L'avantage de la FLD sur la décomposition de PCA est que l'espace de projection FLD maximise la séparabilité de différentes classes.

Les algorithmes directs de LDA (D-LDA) sont proposés [96–98] pour empêcher la perte de l'information discriminante qui se produit quand une décomposition de PCA est appliquée

avant la LDA. Les algorithmes de décomposition par LDA sont généralisés vers des versions de *noyaux*, nommés aussi *analyse discriminante générale* (GDA: General Discriminant Analysis) [99] ou *analyse discriminante par noyau de Fisher* (KFDA: Kernel Fisher Discriminant Analysis) [100]. Avec GDA/KFDA, l'espace original d'entrée est projeté, par mappage non-linéaire, vers un espace de caractéristiques à haute dimension où les différentes classes de visages sont linéairement séparées [99,101].

La puissance de la discrimination de ces algorithmes dépend fortement du nombre d'échantillons par classe, qui est inclus dans l'ensemble de données. En conséquence, la capacité de la généralisation de ces méthodes s'appuie fortement sur les variations de pose et d'éclairage enregistré pour chaque objet [102,103].

L'*analyse en composantes indépendantes* (ICA: Independent Component Analysis) est une autre méthode basée sur le sous-espace qui vise à trouver des projections linéaires (des vecteurs à sous-espace) minimisant l'indépendance statistique entre ses composantes [104]. ICA définit un modèle génératif pour les données multi-variables observées, qui sont typiquement représentées comme une large base de données d'échantillons. Dans ce modèle, les variables de données sont considérées comme des mélanges linéaires de variables latentes inconnues, le système de mélange est aussi inconnu. Les variables latentes sont supposées non-gaussiennes et mutuellement indépendantes. Elles sont nommées les composantes indépendantes de données observées. Ces composantes indépendantes, également nommées *sources* ou *facteurs*, peuvent être trouvées par l'ICA. De même que la PCA, l'ICA donne une projection linéaire $R^N \rightarrow R^M$ mais avec des propriétés différentes. En pratique, l'ICA est destinée à décomposer le signal d'entrée (les images de visages) en une combinaison linéaire de sources indépendantes. L'hypothèse de base est:

$$x^T = As^T \quad (2.5)$$

Où A est la matrice de mélange inconnue. Avec les algorithmes type ICA, nous essayons de trouver la matrice de séparation W tel que:

$$Wx^T = WAs^T \quad (2.6)$$

La décomposition de l'ICA peut être appliquée au problème de la reconnaissance de visages de deux façons différentes:

- Les images de visages sont les variables du système, les images basiques indépendantes sont les rangées de la matrice S et les pixels sont les observations. Par conséquent, les vecteurs basiques obtenus représentent des caractéristiques locales pour les images de visages.
- Les sources sont des coefficients indépendants (les valeurs propres) calculés d'un SVD appliqué à l'ensemble des données, les images basiques sont les colonnes de la matrice de mélange A et les variables sont les pixels. Les vecteurs basiques représentent les propriétés globales pour les images de visages et chaque vecteur représente un mode de variation du visage entier. Les vecteurs basiques calculés ressemblent plus étroitement aux vecteurs de la décomposition de PCA.

Comme montré dans la figure 2.5, où les vecteurs basiques de la PCA sont orthogonaux et alignés sur la direction de la variabilité maximale des données, les vecteurs basiques résultants de l'ICA sont non-orthogonaux mais alignés le long des deux directions de la variabilité maximale des données. Afin de mieux présenter les propriétés locales du visage, l'ICA est également appliquée sur des représentations de visages basées sur la décomposition de *Gabor* [105]. Une version non-linéaire de l'ICA utilisant les fonctions noyaux (K-ICA) a été proposée par *Bach et al.* [106] (2002).

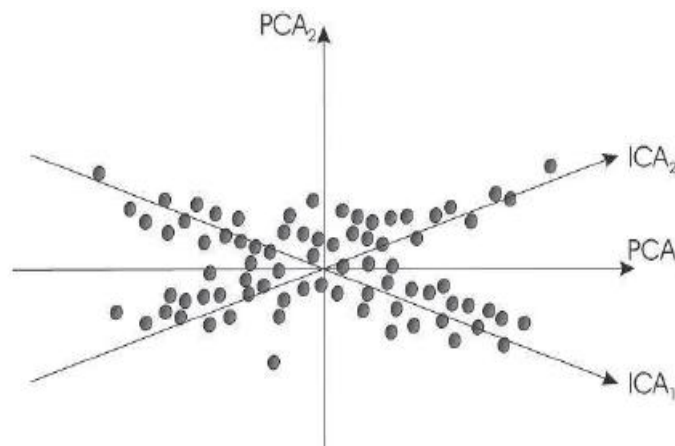


Figure 2. 5 : Composantes principales définies par la PCA (La décomposition de *Karunen Lowe*) et par l'analyse en composante indépendante (ICA).

Toutes les méthodes basées sur les sous-espaces sont très populaires pour mettre en œuvre des systèmes de reconnaissance de visages. Néanmoins, elles sont toutes très sensibles aux changements de l'arrière plan et aux erreurs d'alignement. Pour cette raison, le visage doit être, dans la plupart du temps, coupé manuellement à partir de l'image. D'autre part, puisque

l'ensemble de données est traité comme une seule matrice, la cohérence géométrique doit être imposée pour toutes les instances de visages. En conséquence, toutes les images faciales doivent être soigneusement alignées dans un cadre de référence commun. Une petite erreur en alignement du visage peut induire de très grandes erreurs dans la classification de visages.

b. Caractéristiques faciales et les méthodes basées sur les points de repère (locale / géométrique)

L'attention et les fixations prévues jouent un rôle crucial dans la reconnaissance de visages par les humains. Les processus attentifs, à leur tour, sont généralement guidés par des caractéristiques saillantes qui peuvent être localisées par le calcul d'une carte de saillance dans l'espace considéré. Les mêmes points saillants peuvent fournir des informations utiles aux algorithmes de reconnaissance de visages. En fait, les régions du visage sur l'image ne donnent pas la même quantité d'informations. Par exemple, le front et les joues ont des structures très simples et moins de motifs distinctifs par rapport aux yeux ou nez.

Les points saillants sur le visage sont utilisés dans l'enregistrement des images de visages, dans la normalisation des expressions et pour effectuer une reconnaissance basée sur la distribution géométrique et le motif au niveau de gris à des positions repérées. Même si les études approfondies rapportées dans la craniologie définissent avec précision un ensemble riche de repères faciaux, il n'y a aucun ensemble universel de points de repère accepté pour la reconnaissance. *Bookstein* (1989) [107] a défini les points de repère en tant que « des points sous une forme pour laquelle les contreparties biologiques sont objectivement significatives et reproductibles qui existent dans toutes les autres formes dans un ensemble de données ». Les points de repère les plus fréquemment utilisés sur le visage sont: le bout du nez, les bouts des yeux, les bouts sur les coins de la bouche, le centre de l'iris, le bout du menton, les narines, les sourcils et la nasale. Il est à noter que dans la littérature, les régions discriminantes du visage, telles que les yeux ou la bouche, sont également référées en tant que "*caractéristiques faciales*". Cette terminologie engendre parfois une ambiguïté. En fait, le plus souvent le terme "*caractéristique*" est utilisé dans le domaine de la reconnaissance des formes pour définir une représentation spécifique extraite à partir d'un motif au niveau de gris. Par exemple, les vecteurs d'Eigen-faces sont également nommés "*caractéristiques*", ainsi les représentations numériques obtenues par le multi-canal de *Gabor* appliqué sur une image au niveau de gris sont nommées "*caractéristiques*". Pour cette raison, les motifs extraits à partir de localisations

spécifiques et discriminantes sur l'image du visage sont nommés ici "*points de repère*" au lieu de "*caractéristiques*".

Dans les méthodes basées sur la géométrie, la distribution des points de repère sur le visage est utilisée sous la forme de règles heuristiques qui impliquent des angles, des distances et des régions [108–110]. Dans les méthodes basées sur la structure, la géométrie est incorporée dans un modèle complet de structure. Par exemple, dans l'approche de *correspondance par graphe de tas élastique* (EBGM: Elastic Bunch Graph-Matching), un graphe modélise les positions relatives aux points de repère, où chaque nœud représente un point sur le visage et les arcs sont pondérés en fonction des distances prévues du repère (Figure 2.6). Pour chaque nœud, un ensemble de modèles est employé pour évaluer la similarité locale de la caractéristique [111]. Quoique les déformations possibles dépendent souvent des points de repère (ex., les coins de la bouche se déforment beaucoup plus que le bout du nez), l'information spécifique du repère peut être incorporée dans le modèle structural [112]. Comme l'ensemble de contraintes optimisées conjointement est agrandi, le système fonctionne plus fréquemment avec les problèmes de la convergence et les optimums locaux, qui à son tour rend une bonne – et souvent manuelle – initialisation nécessaire.

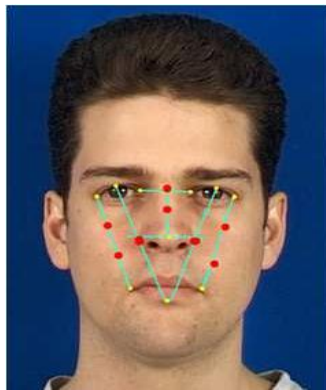


Figure 2. 6 : Extraction des points de repère par l'algorithme EBGM.

Plusieurs approches sont proposées pour dériver des représentations de visages à partir d'un certain nombre de composantes ou de sous-images faciales. *Moghaddam et al.* (1994) [113] ont proposé une version de PCA basée sur des composantes dont le sous-espace du visage est composé d'un certain nombre de sous-espaces construits à partir d'images partielles des images faciales originales. Le point de repère choisi se trouve entre les yeux et la bouche. *Tistarelli* (1995) [114] a proposé une méthode basée sur l'extraction des repères faciaux qui sont ré-échantillonnés par l'application d'un mappage de type *log-polar*. La reconnaissance

actuelle est exécutée par l'application d'une corrélation croisée et normalisée entre deux représentations de visages. La valeur de la corrélation détermine la similarité entre les deux représentations et, par conséquent, les deux objets. La valeur de la corrélation est utilisée ensuite comme un score dans la classification.

La correspondance par graphe élastique (EGM: Elastic Graph Matching) est une autre classe populaire dans la reconnaissance de visages qui se base sur les points de repère. EGM est une implémentation pratique de l'architecture dynamique des arcs pour la reconnaissance d'objets [115,116]. Avec EGM, le graphe d'objet référentiel est créé en superposant un graphe clairsemé, élastique et rectangulaire sur l'image de l'objet et en calculant la réponse d'une banque d'ondelette de *Gabor* à chaque nœud du graphe. La valeur accumulée à chaque nœud représente un *jet*. Le processus de la correspondance par graphe est mis en œuvre par l'optimisation stochastique d'une fonction de coût qui tient en compte les similarités des jets et la déformation du nœud. Une procédure d'optimisation en deux étapes grossières suffit pour la minimisation d'une telle fonction de coût :

$$s_{\phi}(J^0, J^1) = \frac{\sum_j a_j^0 a_j^1 \cos(\phi_j^0 - \phi_j^1) - \bar{d} \bar{k}_j}{\sqrt{\sum_j (a_j^0)^2 \sum_j (a_j^1)^2}} \quad (2.7)$$

Où J^0 et J^1 sont les deux ensembles de jets à comparer, (a_j^0, ϕ_j^0) et (a_j^1, ϕ_j^1) représentent les coefficients jets (amplitude et phase) pour les deux ensembles, \bar{d} est le déplacement entre les deux ensembles et \bar{k}_j est le vecteur d'onde ou l'ensemble des coefficients noyaux.

La correspondance par graphe de tas élastique (EBGM: Elastic Bunch Graph-Matching) est une variation de l'EGM. Dans la structure du graphe de tas, un ensemble de jets est calculé à chaque nœud pour différentes instances du même visage (ex., avec la bouche et les yeux ouverts ou fermés) [117,118]. De cette façon, la représentation du graphe de tas peut gérer plusieurs variations dans l'apparence du visage.

La correspondance par graphe élastique morphologique (MEGM: Morphological Elastic Graph Matching) [119–121] est une autre approche qui est similaire à l'EGM. Dans ce cas, les caractéristiques de *Gabor* sont remplacées par des caractéristiques morphologiques multi-échelles obtenues par un filtrage de dilatation-érosion de l'image faciale [122].

Les techniques discriminantes sont utilisées pour améliorer les performances d'identification et vérification de toutes ces approches. L'application de techniques

discriminantes linéaires sur les vecteurs de caractéristiques pour la sélection des coefficients les plus discriminants a été proposée dans [118,119,123]. Plusieurs arrangements qui visent à pondérer les nœuds du graphe en fonction de leur puissance discriminative ont été proposés [119,123–125]. Dans [123], la sélection des coefficients de pondération est basée sur une fonction non-linéaire qui dépend à un ensemble restreint de paramètres. Ces derniers sont déterminés sur l'ensemble d'apprentissage en maximisant un critère par l'utilisation de la méthode du *simplex*. Dans [119], l'ensemble des coefficients de pondération des nœuds n'est pas calculé par certains critères d'optimisation, mais par l'utilisation des statistiques du premier et deuxième ordre sur les valeurs de similarité des nœuds. Une approche *Bayésienne* pour déterminer quels sont les nœuds les plus fiables a été utilisée par *Chang et al.* (2003) [123]. Un arrangement plus sophistiqué pour pondérer les nœuds du graphe élastique en construisant une classe modifiée à l'aide des séparateurs à vastes marges (SVM: Support Vector Machines) a été proposée par *Chang et al.* (2005) [125]. Dans ce travail, les auteurs ont également montré que les performances de la vérification de l'EGM peuvent être fortement améliorées par des stratégies de pondération des nœuds appropriés. Le sous-espace des algorithmes de vérification et d'identification de visages considère l'image entière comme un vecteur de caractéristiques, son but consiste à trouver des projections qui optimisent un certain critère défini sur les vecteurs de caractéristiques qui correspondent aux différentes classes.

L'inconvénient principal de toutes les méthodes basées sur les points de repère est celui d'exiger des images faciales parfaitement alignées [126]. C'est-à-dire, toutes les images faciales doivent être alignées afin d'avoir tous les points référentiels (ex., les yeux, le nez et la bouche) représentés à la même position au niveau du vecteur de caractéristiques. A cet effet, les images faciales sont très souvent alignées manuellement et d'ailleurs elles sont mises sous une échelle anisotrope (standard). L'alignement automatique parfait est en général une tâche difficile à évaluer. En revanche, la correspondance par graphe élastique ne nécessite pas un alignement parfait pour bien fonctionner. Le principal inconvénient de l'EGM réside dans le temps requis par l'analyse multi-échelle de l'image faciale et la procédure de correspondance. Il est communément connu que les différences dans l'illumination qui envisagent la reconnaissance du visage présentent l'un des défis majeurs dans cet aspect. La manière dont les ordinateurs conçoivent la géométrie du visage humain est également un problème que les

chercheurs sont appelés à résoudre afin d'augmenter la robustesse et la stabilité d'un système de reconnaissance de visages.

2.4.4 Traitement de l'illumination

La reconnaissance de visages (au moins en 2D) est principalement basée sur l'analyse des propriétés photométriques de la surface du visage. Pour cette raison, les changements de la réflectivité du visage aux variations d'illumination (orientation ou énergie de la source) ne peuvent pas être distingués à cause des déformations dans la forme. Afin de faire face à ce problème, plusieurs méthodes sont étudiées pour (i) normaliser l'illumination du visage (ii) ou calculer la composante de l'illumination indépendante sur la surface du visage. Un aperçu sur les méthodes de la compensation d'illumination pour la reconnaissance de visages peut être trouvé dans la référence [127]. Les méthodes proposées pour la compensation d'illumination tombent dans l'une de ces trois catégories principales:

- **Techniques basées sur l'histogramme adaptatif:** L'image faciale est divisée en sous-régions qui sont analysées au moyen de techniques d'histogrammes adaptatifs pour normaliser le niveau d'intensité de chaque sous-région. Plusieurs approches sont présentées dans les références [128] et [129].
- **Techniques basées sur la ré-illumination:** Le niveau d'intensité de chaque pixel de l'image peut être considéré comme le rapport de deux composantes: l'énergie lumineuse frappant la surface de la peau et la réflectivité réelle de la peau. La réflectivité de la peau est la composante d'illumination indépendante qui peut être, ensuite, calculée comme le rapport entre l'intensité et les valeurs d'illumination. Ces méthodes visent à estimer l'illumination en pixel par pixel et par conséquent la réflectivité de la peau.
- **Synthèse des représentations de l'illumination-invariante:** Comme exemple, le composant *Hue* dans l'espace couleur HSV, est invariant aux ombres. Un autre exemple est l'espace *tubulure* du visage, y compris plusieurs instances des mêmes visages avec des directions variantes d'illumination.

La nature complexe du tissu de la peau humaine, rend la modélisation exacte de la réponse sur la surface du visage une tâche très difficile en illumination. La peau n'est pas un matériel simple; elle est composée de plusieurs couches de tissu ayant différentes propriétés chromatiques et réflectivités lumineuses [130]. Pour cette raison, n'importe quel modèle

dérivé de la réflectivité de la peau peut être seulement une approximation de véritables propriétés de la surface et ne peut être aussi valable que dans certaines hypothèses.

La méthode proposée par *Qing et al.* (2004) [131] appartient à la deuxième catégorie, elle est basée sur des images harmoniques. Comme les visages humains partagent une forme semblable et la surface du visage est quasi-constante, les neuf composantes, de basse fréquence, de l'éclairage sont estimées à partir des images d'entrées. L'image du visage est normalisée par une procédure de ré-éclairage en fonction du rapport de l'illumination de l'image. Les auteurs proposent d'étalonner l'image du visage en entrée à une illumination canonique. Les résultats expérimentaux montrent une amélioration significative dans les performances de la reconnaissance de visages.

Dans le travail de *Gross et Brajovic* (2003) [132], le champ de luminance est calculé au moyen d'une diffusion anisotrope dérivée de la minimisation de la fonction:

$$F(L) = \int \int_{\omega} p(x, y)(L(x, y) - I(x, y))^2 dx dy + \lambda \int \int_{\omega} (L_x^2 + L_y^2) dx dy \quad (2.8)$$

Où la première intégrale contraint la luminance $L(x, y)$ à être proche de l'intensité des pixels capturés de l'image $I(x, y)$, la deuxième intégrale impose le lissage du champ de luminance récupéré, commandé par le paramètre λ , et $p(x, y)$ détermine le niveau d'anisotropie du processus de la diffusion. La fonction peut être résolue par les multiplicateurs de *Lagrange* dans le but de déterminer la valeur de $L(x, y)$. $p(x, y)$ varie en fonction de la luminance de l'image, elle est définie en tant que la réciproque de la fonction de contraste de *Weber*:

$$\frac{p(i, j) = |I_i I_j|}{\min(I_i I_j)} \quad (2.9)$$

L'approche proposée par *Wang et al.* (2003) [133] entre dans la troisième catégorie. Un sous-espace d'illumination du visage est conçu sur la base de trois ou plusieurs images de formation illuminées par des lumières non-coplanaires. L'éclairage de n'importe quelle image faciale est représenté comme un point dans ce sous-espace. Ce travail offre principalement un cadre très général pour l'analyse et la modélisation des images faciales dans des conditions diverses d'éclairage. Le concept de l'espace d'illumination du visage est ainsi introduit comme un modèle de visage général pour la modélisation de l'espace du visage sous

conditions variantes d'éclairage. Dans un système pratique, basé sur l'analyse du sous-espace, l'illumination et la pose sont deux problèmes auxquels nous devons faire face simultanément.

Dans l'approche proposée par *Qing et al.* (2004) [134], les statistiques de la dérivée des images irradiantes (log) du visage humain sont analysées. Une mesure de distance insensible à la lumière est définie en fonction de l'opérateur minimum des dérivées de deux images. La mesure proposée pour récupérer la composante de réflectivité est comparée à l'opérateur médian proposée par *Weiss* (2001) [135]. Quand les tests sont sous illuminations variantes, les expériences de la reconnaissance de visages sur la base d'images CMU-PIE montrent que la mesure proposée est bien meilleure que la corrélation d'intensité de l'image et mieux que la distance euclidienne de la dérivée de l'image *log* utilisée par *Turk et Pentland* (1991) [89].

2.4.5 Représentation numérique des visages humains

Les méthodes holistiques pour l'identification de visages nécessitent un grand ensemble de données d'apprentissage (statistiquement significatives) pour établir des vecteurs basiques qui déterminent l'espace de faible résolution. Les capacités de la généralisation de ces méthodes sont testées dans certaines mesures, qui sont peu claires. Jusqu'à présent les tests sont effectués sur des bases d'images ayant des tailles limitées. Même la base d'images FRGC [136] comporte seulement quelques milliers d'objets. L'escalade aux grandes bases de données, y compris des centaines de milliers de personnes, quand c'est possible, rend le problème très difficile à analyser numériquement. Le management d'identité par ces représentations de visages nécessite d'être capable de distinguer chaque individu à travers un espace de caractéristiques unique, mais c'est difficilement garanti. Les meilleures méthodes performantes en reconnaissance de visages, basées sur le traitement holistique dans les conditions réelles, atteignent un taux d'équivalence des erreurs (EER) autour de 1%. Cela correspond à 100 objets incorrectement classés sur une base d'images de 10 000 individus ou 1000 objets sur une base d'images de 100 000 individus. La taille du modèle de caractéristiques dépend de la dimensionnalité de l'espace de représentation, c'est-à-dire du nombre de vecteurs basiques sélectionné pour la représentation de la base de données. Cette valeur dépend de la population d'objets, de la variabilité sur l'apparence du visage (pose, expression, éclairage, etc.), du nombre de classes et de la puissance discriminative. Par conséquent, quand il s'agit de faire face aux nombreuses variations sur l'apparence du visage, par exemple le vieillissement et la taille du sous-espace, la représentation peut devenir

indéfiniment grande. L'un des avantages des méthodes basées sur les caractéristiques est la forte dépendance du signal d'entrée plutôt que la population des objets analysés. Pour cette raison, l'identité de l'objet est représentée exclusivement par l'information liée aux données capturées de chaque objet. Le modèle du visage résultant peut être très compact car il est limité à la structure géométrique des caractéristiques avec les valeurs associées. Cela permet de faire face à beaucoup de variations, y compris le vieillissement, sans affecter la taille de la représentation.

Les études réalisées par des expériences neuropsychologiques concernant la perception du visage permettent de déduire également quelques indications sur la complexité de la représentation de visages dans le cerveau humain. *Leveroni et al.* (2000) [137] ont défini 31 régions dans le cerveau humain, principalement liées à la sulcature temporelle supérieure (Figure 2.1) et au striatum ventral, qui sont activés lors de la perception du visage. Le volume total des zones activées se résume à 21,2 ml. Etant donné une masse totale de 1400 ml pour le cortex du cerveau, avec un total approximatif de 100×10^9 neurones, nous nous retrouvons à un total d'environ de $1,5 \times 10^9$ neurones. Comptant en moyenne de 12 000 synapses par neurone dans le cortex visuel, les zones activées durant la perception du visage se somment à 18×10^{12} synapses. Etant limité la représentation d'un bit d'information par synapse, les informations transportées par le réseau de neurones sont plus ou moins équivalentes à $2,3 \times 10^{12}$ octets. Si nous considérons un scénario simple où un objet apprend typiquement à identifier 1000 visages, la quantité d'informations qui peut être potentiellement stockée dans un tel réseau est égale à un flux vidéo d'environ 70 secondes avec une résolution de 1 Mega-Pixel par image. Même si l'architecture neuronale soumet la perception de visages dans le cerveau humain, elle est aussi capable de traiter les changements complexes tels que les expressions faciales et le vieillissement, cette comparaison simple nous permet toujours de comprendre la complexité et la quantité d'informations impliquées dans la perception de visages et la reconnaissance des individus en mouvements

2.4.6 Reconnaissance de visages à distance

Le visage de l'être humain est un point de repère très distinctif dans le corps humain qui est bien approprié pour effectuer une reconnaissance à distance. La plupart des dispositifs faciaux visibles peuvent être capturés même si la résolution spatiale de la caméra est faible, et

donc il n'est pas nécessaire de capturer des images à une distance très proche. Afin d'acquérir des images de visages à distance, trois processus sont impliqués:

- a. Segmentation du corps mobile,
- b. Détection et suivi de la tête, et
- c. Localisation du visage.

Les trois processus nécessitent un échantillonnage approprié de la séquence dans l'espace et le temps pour isoler le contenu de l'information qui sera pertinente pour un traitement ultérieur. La détection et la segmentation du corps humain sont récemment étudiées dans le contexte de la surveillance visuelle. Plusieurs méthodes proposées sont basées sur des techniques de soustraction de l'arrière plan telles que *les modèles de mélange Gaussien*. Le raisonnement soutenant toutes ces méthodes, est la définition d'un mécanisme d'apprentissage approprié capable de capturer les fréquences de caractéristiques spatio-temporelles d'un corps en mouvement. D'autres considérations et détails sur la méthodologie à appliquer dans la capture des images de visages sont rapportés dans la référence [138], spécifiquement consacrée à l'acquisition des images faciales à distance.

2.5 Défis ouverts et futures applications

Comme dans la plupart des applications biométriques, les variations provoquées sur l'apparence par les conditions non-contraintes sont toujours des défis pour la reconnaissance de visages. Parmi ces défis, nous citons:

2.5.1 Reconnaissance de visages et les occlusions

Dans un environnement réel, le visage peut être capturé dans une pose arbitraire et sans aucune coopération de l'utilisateur, donc il est très probable que l'image contient seulement un visage partiel. Les visages sont facilement occlus par des accessoires faciaux (ex., lunettes de soleil, foulard, chapeau et voile), des objets devant le visage (ex., main, nourriture et téléphone mobile), l'illumination extrême (ex., l'ombre), l'auto-occlusion (ex., pose non-frontale) ou la mauvaise qualité de l'image (ex., le flou). Dans l'identification de visages en criminalistique, par exemple, il est nécessaire de trouver un suspect dans la foule, par la correspondance d'un visage partiellement occlus sur une base d'images enregistrée. La difficulté dans la reconnaissance de visages occlus est double; premièrement, l'occlusion déforme les caractéristiques discriminatives faciales et augmente la distance entre deux

images d'un même objet dans l'espace de caractéristiques. Les variations intra-classes sont plus grandes que les variations inter-classes, qui ont pour conséquence une mauvaise performance de reconnaissance. Deuxièmement, quand les points de repère faciaux sont occlus, les grandes erreurs d'alignement se produisent habituellement et dégradent les taux de reconnaissance [139].

Une idée intuitive pour gérer l'occlusion dans la reconnaissance automatique de visages consiste d'abord à détecter la région occluse et ensuite effectuer une reconnaissance en utilisant uniquement les parties non-occluses. Cependant, les types d'occlusions sont imprévisibles dans les scénarios pratiques. L'endroit, la taille et la forme de l'occlusion sont inconnus, ce qui augmente la difficulté en segmentation de la région occluse dans l'image du visage. Une manière plus pratique consiste à effectuer la reconnaissance avec la présence de l'occlusion. Il existe deux grandes catégories d'approches principales dans cette direction:

Le premier type réside dans *les approches basées sur la reconstruction* qui traitent la reconnaissance de visages occlus comme un problème de reconstruction [140–147]. La classification basée sur *la représentation clairsemée* (SR: Sparse Représentation) proposée par *Wright et al. (2009)* [145] est un excellent exemple représentatif; une image propre est reconstruite à partir d'une image de sonde occluse par une combinaison linéaire d'images de galerie et les vecteurs basiques d'un dictionnaire d'occlusions. Ensuite, l'image occluse est assignée à la classe correspondante avec l'erreur minimale de la reconstruction.

Le deuxième type réside dans *les approches basées sur la correspondance locale*. Les caractéristiques faciales sont extraites à partir de régions locales dans un visage, par exemple, les pièces chevauchantes et non-chevauchantes d'une image, de sorte que les parties affectées et non-affectées du visage peuvent être analysées en isolation. Afin de minimiser les erreurs de correspondance à cause de parties isolées, différentes stratégies sont effectuées telles que: la pondération [148], le gauchissement [149,150], le vote [151], l'apprentissage avec un espace local [83,152], ou l'apprentissage par représentation clairsemée multitâche [153].

Klontz et Jain (2013) [154] ont conduit une étude de cas, qui a utilisé les photographies de deux suspects dans les attentas du marathon de *Boston* dans le but de faire une correspondance avec un ensemble de Mugshots (Figure 2.3). Les photographies des suspects libérées par le FBI ont été capturées dans un environnement non-contrôlé et leurs visages ont été partiellement occlus par des lunettes de soleil et chapeaux. L'étude a montré que le

système commercial actuel de reconnaissance de visages a eu le potentiel notable pour aider les forces légales. Mais la précision de la correspondance n'était pas élevée et plus de progrès doit être fait pour augmenter l'utilité en images de visages non-contraintes.

2.5.2 Reconnaissance de visages hétérogènes

La reconnaissance de visages hétérogènes implique la correspondance de deux images de visages à partir de modalités d'imageries alternatives. Ceci est très pratique dans les scénarios légaux. Par exemple, dans les émeutes de *Londres* en 2011, un système de reconnaissance de visages a été utilisé par la police pour aider à trouver les suspects impliqués dans les troubles. Les images des suspects sont venues de diverses sources, par exemple, des images fixes capturées à partir de caméras en circuit fermé, des images recueillies par des officiers, des images prises par les hélicoptères de la police ou des images saisies par des membres du public. Ces images sont généralement de diverses sources de différentes modalités. En outre, dans certaines situations extrêmes, seulement une modalité particulière d'une image de visage est disponible. Par exemple, dans les environnements de nuit, *l'imagerie infrarouge* (Figure 2.7.c) peut être la seule modalité pour acquérir une image utile d'un visage de suspect. Mais les Mugshots enregistrés par la police sont des images de bande visible. Un autre exemple réside dans la correspondance par des photographies de type *Sketch* (Figure 2.7.b); quand aucune image d'un suspect n'est disponible, un *sketch légal* est souvent généré selon la description d'un témoin oculaire. La correspondance par les sketches contre les photographies de visages est très importante dans les enquêtes légales.

Il existe actuellement trois catégories d'approches dans la reconnaissance de visages hétérogènes. La première est *la méthode basée sur les caractéristiques* [155] qui représente les images de visages avec les caractéristiques distinctives qui sont invariables dans les différentes modalités de l'image. La deuxième est *la méthode basée sur la synthèse* [156] qui convertit une image de visage en une seule modalité (ex., sketch) par rapport aux autres (ex., photographie). La troisième est *la méthode basée sur le prototype* [157] qui réduit l'écart entre les deux modalités en utilisant un prototype comme un pont. La correspondance par visages 2D-3D est une future direction de recherche car le visage peut être représenté par des caractéristiques hétérogènes dans les modalités 3D et 2D dans le cas du monde réel.

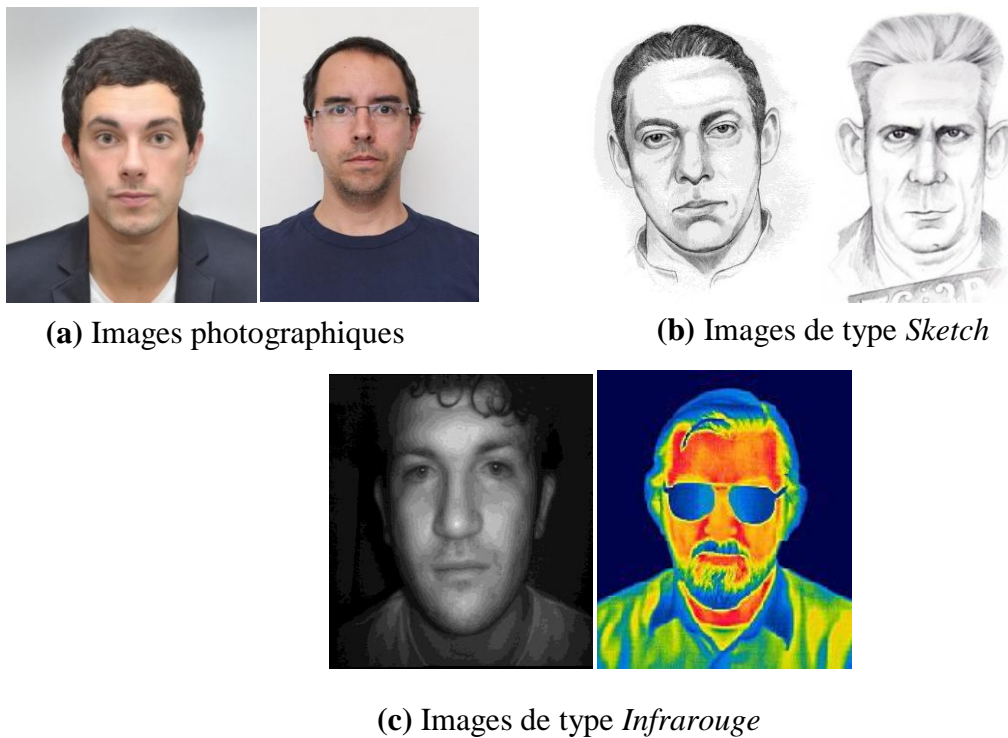


Figure 2. 7 : Quelques modalités d'imageries qui démontrent des hétérogénéités alternatives.

2.5.3 Reconnaissance de visages et le vieillissement

Le vieillissement du visage est un processus complexe qui affecte à la fois la forme et la texture (ex., le ton ou les rides de la peau) d'un visage. Le scénario typique pour l'application des systèmes de reconnaissance de visages contre l'effet du vieillissement, est de détecter si une personne particulière est présente dans une base de données enregistrée précédemment (ex., l'identification des enfants disparus et le contrôle des suspects sur une liste de surveillance). Comme l'âge entre une image de requête et une autre de référence du même objet augmente; donc la précision d'un système de reconnaissance diminue généralement.

En reconnaissance faciale automatique, l'effet du vieillissement sur les visages humains a été étudié dans deux directions: (i) le développement de *techniques d'estimation d'âge* pour classifier les images faciales en fonction de l'âge [158,159] et (ii) le développement de *systèmes de vieillissement robustes* pour effectuer la reconnaissance. Dans un premier temps, les chercheurs ont essayé de simuler les effets du vieillissement en développant une fonction de vieillissement et puis en effectuant une estimation automatique de l'âge par la technique de *Lanitis et al.* (2002) [160]. Mais la modélisation complexe de la forme ou les variations de la texture d'un visage en vieillissement est une tâche très difficile. Aujourd'hui, les chercheurs

proposent un modèle de vieillissement génératif [161], qui apprend un modèle paramétrique de vieillissement dans le domaine 3D afin de générer des images synthétiques et de réduire l'écart d'âge entre la requête et les images de référence. L'aspect le plus difficile en reconnaissance de visages à travers le vieillissement est qu'il doit aborder toutes les autres variations non-contraintes. La figure 2.8 montre des échantillons de visages de deux personnes dans le vieillissement. La pose, l'expression faciale, le changement de l'illumination et l'occlusion peuvent se produire lorsque les images sont prises à des années d'écart.

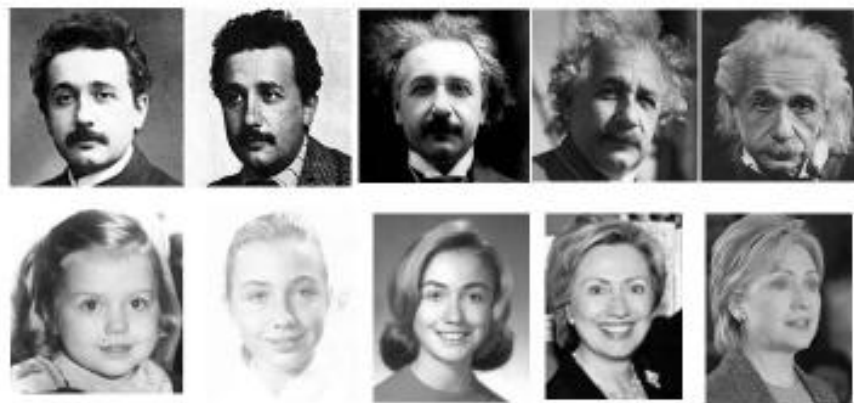


Figure 2. 8 : Echantillons de visages de deux personnes dans le vieillissement.

2.6 Conclusion

L'étude de la reconnaissance de visages humains par des psychologues et neurologistes a été dirigée en parallèle avec le développement de technologies de reconnaissance automatique par des informaticiens et ingénieurs. Dans les deux cas, des étapes analogues dans l'acquisition de données, le traitement de l'image et la formation des représentations peuvent soutenir les tâches complexes et variées que nous accomplissons avec les visages. Ces processus peuvent être compris et comparés dans le contexte de leurs implémentations neuronales et informatiques. A travers ce chapitre, nous avons présenté les éléments essentiels sur la reconnaissance de visages par les humains et les machines, en prenant une perspective qui englobe des approches psychologiques, neurologiques et informatiques. Nous avons donc présenté d'une part les méthodes et les techniques utilisées en neurobiologie de la reconnaissance de visages, le rôle de l'attention visuelle et la nature qui émerge les représentations. D'autre part, nous avons discuté des technologies de reconnaissance de

visages et des stratégies qu'elles utilisent pour surmonter les défis pour un fonctionnement robuste sous différents paramètres de visualisation. Finalement, nous avons terminé le chapitre par la présentation de quelques défis ouverts, des solutions récentes proposées et les futures recherches dans ce domaine

Chapitre 3

Reconnaissance Biométrique par Oreilles

3.1 Introduction

La possibilité d'identifier des personnes par utilisation de la forme extérieure de l'oreille a été découverte, pour la première fois, par le criminologiste français *Alphonse Bertillon* et confirmée par l'officier de police américain *Alfred Iannarelli*. Ce dernier a proposé le premier système d'identification automatique par oreilles, basé uniquement sur sept caractéristiques. La structure détaillée de l'oreille est non seulement unique, mais aussi stable, car l'apparence de l'oreille ne change pas durant la vie de l'être humain. En plus, l'acquisition des images de l'oreille n'exige pas, nécessairement, la coopération de la personne, elle est considérée par la plupart des chercheurs comme une modalité non-intrusive. En raison de ces qualités, l'intérêt des systèmes de reconnaissance par oreilles a été développé d'une manière très significative au cours de ces dernières années.

Dans ce chapitre, nous présentons une étude anatomique et embryologique de l'oreille humaine dans le but de connaître les emplacements sur lesquels nous pouvons atteindre une grande variation interindividuelle. Ensuite, nous récapitulons les approches de détection et reconnaissance automatique d'oreilles. Finalement, nous présentons quelques perspectives sur les problèmes non-résolus et les futures recherches dans ce domaine, spécialement dans le contexte de la surveillance intelligente et la criminalistique.

3.2 Caractéristiques de l'oreille : Approche biologique

La formation de l'oreille dans l'embryon humain est généralement prise comme un développement individuel de ses composantes séparées. L'identification des composantes qui forment la structure complexe de l'oreille humaine, est considérée comme un souci principal dans les études embryologiques de l'oreille. *Pourquoi nous intéressons nous, dans cette thèse à l'embryologie de l'oreille ?* En effet, la prémisse des structures locales et indépendantes dans le pavillon auriculaire revêt un grand intérêt dans notre approche d'identification

proposée. Avant de rentrer dans les détails de l'embryologie, nous commençons d'abord par la présentation de la terminologie anatomique de l'oreille.

3.2.1 Anatomie de l'oreille humaine

La figure 3.1 présente une oreille avec les diverses parties. La composante la plus saillante de l'oreille réside dans sa jante externe, appelée *Hélix*, qui se fusionne au fond avec le *Lobule*. L'*Anthélix* est le bord arrondi de la *Conque* qui est presque parallèle à l'*Hélix*. Il bifurque dans deux branches au dessus, formants le *Cru Supérieur* et le *Cru Inférieur* de l'*Anthélix*. La *Conque* est une cavité qui a une forme de coquille, entourant le canal auditif externe et se fusionne avec l'*Incisure*. L'*Incisure* a deux bosses sur chaque côté qui sont le *Tragus* et l'*Antitragus*. Le canal auditif extérieur est généralement masqué par le *Tragus*. La *Conque* est divisée en deux parties par le *Cru de l'Hélix* (la partie horizontale de l'*Hélix*), qui forme une arête transversale sur la surface de la *Conque*. La partie de l'*Hélix* directement attachée au *Cru de l'Hélix* est appelée *Hélix ascendant* ou *Antérieure* [162,163]. La figure 3.2 présente une collection d'images d'oreilles qui illustre des variations interindividuelles, ainsi que les parties de base qui sont distinguables dans la plupart des images d'oreilles.

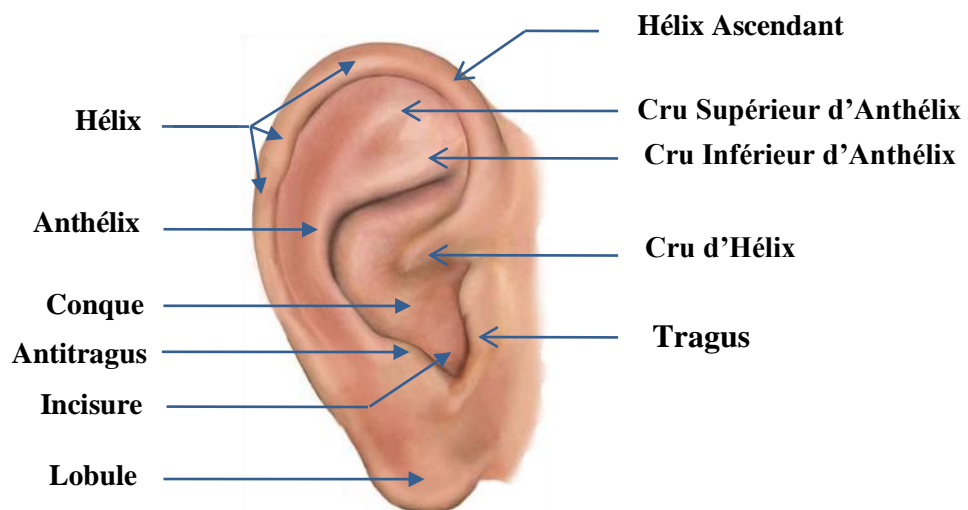


Figure 3. 1 : Anatomie de l'oreille humaine [164].



Figure 3. 2 : Présentation des variations interindividuelles entre les composantes de l'oreille.

3.2.2 Embryologie de l'oreille

L'aspect initial de la forme extérieure de l'oreille dans l'embryon humain apparaît dans une forme de six tertres (*Hillocks*) différents à partir de la sixième semaine de la vie embryonnaire [165]. Ces *hillocks* progressent et se fusionnent jusqu'à l'apparition de la forme finale du pavillon auriculaire. La figure 3.3.a montre un embryon âgé de six semaines avec ses *hillocks* auriculaires. Ces *hillocks* sont numérotés dans le schéma de la figure 3.3.b. Comme illustré dans la figure 3.3.c, la partie extérieure de l'oreille provient d'un tissu de la première et la deuxième voûte branchiale, appelées les voûtes mandibulaires (*Mandibular Hillocks*) et hyoïdes (*Hyoid Hillocks*). Les six *hillocks* auriculaires se composent de trois *hillocks* mandibulaires et trois *hillocks* hyoïdes. Les voûtes mandibulaires et hyoïdes sont séparées par la première fissure branchiale qui provoque le canal auditif extérieur.

En littérature, plusieurs hypothèses concernant la formation extérieure de l'oreille sont préoccupées à identifier la maturation de chacun de ces six *hillocks*, ces derniers ont été observés la première fois par *His* en 1882 [165]. Aujourd'hui, il y a une grande différence d'opinions sur une embryologie précise de la forme extérieure de l'oreille. La figure 3.4 récapitule les arrangements suggérés par différents auteurs, qui répartissent les différents *hillocks* en plusieurs combinaisons liées à une formation exclusive de l'oreille. Les disparités principales dans ces arrangements semblent être dans l'arrangement de *Wood-Jones* [169] où il a assigné trois *hillocks* au tragus, et dans celui de *Karmody* [170] qui a assigné l'origine de l'hélix au *hillock* 3.

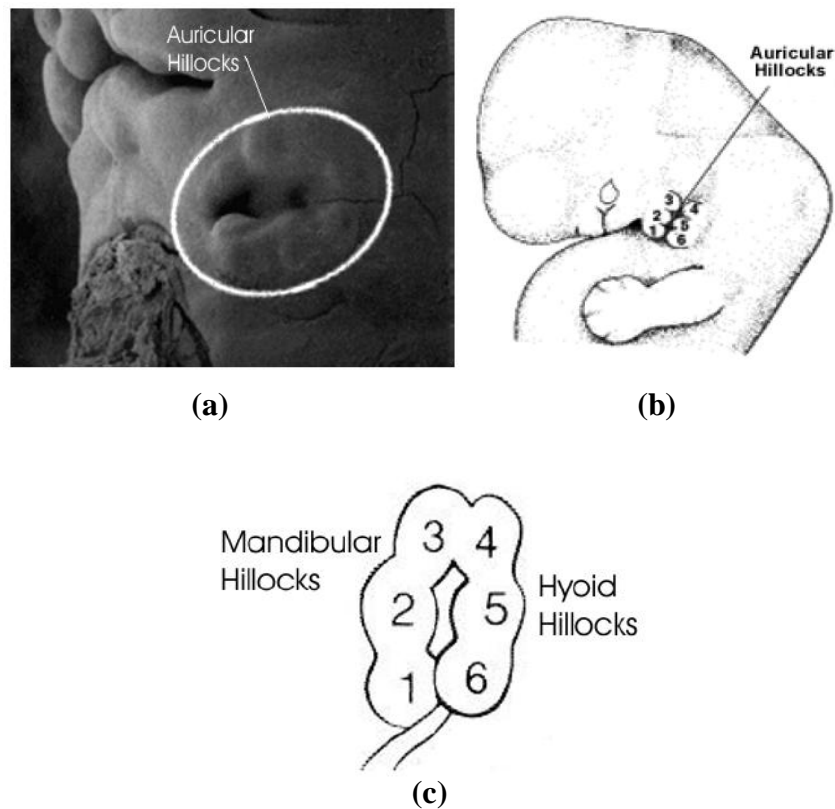


Figure 3. 3 : (a) Image d'un micrographe électronique d'un embryon humain âgé de six semaines [166]. (b) Six hillocks auriculaires sont bien illustrés dans ce schéma [167] (c) Voûtes mandibulaires et hyoïdes [168].

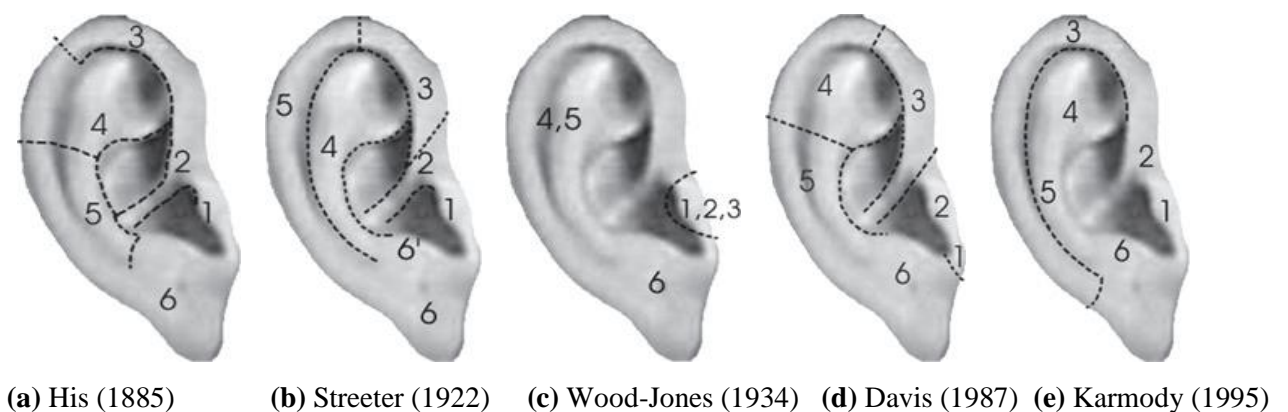


Figure 3. 4 : Différentes positions finales des hillocks, suggérées par différents auteurs [168,169,171].

Considération: Bien que les embryologistes présument et argumentent l'identification exacte du hillock qui forme un élément spécifique du pavillon auriculaire, notre souci consiste, tout simplement, à identifier l'ensemble des composantes autonomes et aussi leurs emplacements sur lesquels nous pouvons atteindre une grande variation interindividuelle.

Contribution: L'étude des malformations de la forme extérieure de l'oreille est considérée comme l'une des principales approches pour comprendre leur embryologie. Cela est particulièrement bénéfique à notre recherche puisqu'elle fournit également des conseils sur les variations possibles de la structure extérieure de l'oreille.

3.2.3 Anomalies extérieures de l'oreille et les composantes sensibles aux variations

Streeter [165] a donné des argumentations sur le développement individuel des hillocks auriculaires; il a suggéré que la forme extérieure de l'oreille hérite son existence comme une structure intacte et continue, qui se mature vers une forme auriculaire finale. Cependant, il y a une diversité de défauts qui perturbe la continuité douce du pavillon auriculaire. Celle-ci peut être mieux décrite comme un échec de fusion ou manque d'alignement correct de ses divers hillocks, ce qui influe négativement sur le rôle des structures séparées dans la formation définitive du pavillon auriculaire [168,171].

Quelques autres malformations peuvent être décrites en tant que croissance excessive ou sous-développement au-dessous des seuils de la normalité. En effet, l'emplacement d'une telle anomalie se trouve, généralement, lorsqu'une variation considérable est présentée; il est peu probable d'observer une anomalie dans les endroits de structures constantes. Comme exemple, le cru de l'hélix est considéré par *Streeter* comme l'une des moindres parties variables du pavillon auriculaire; la présente partie n'est pas un emplacement courant pour des anomalies. Une discussion détaillée sur les anomalies extérieures de l'oreille, quoiqu'elle soit fructueuse pour comprendre sa structure dépasse la portée de notre étude. Ici, Nous mentionnons uniquement les variations des composantes les plus importantes et les plus sensibles qui doivent être prises en considération durant le développement des applications biométriques à base d'oreilles.

La jante externe de l'oreille, l'*hélix*, est généralement attribuée à deux ou trois hillocks embryonnaires. Selon *Streeter* [165] et *Sulik* [166], l'*hélix ascendant*, la partie de l'hélix immédiatement attachée au cru de l'hélix, est assigné à un hillock individuel. Une anomalie nommée: *la branche de l'oreille* est le produit d'absence de l'hélix ascendant, tandis que le reste des autres parties ont leurs formes normales [171]. Deux autres défauts présentant des lacunes remarquables qui séparent l'hélix ascendant au reste des composantes hélicoïdales sur

chaque côté [168,172]. L'hélix ascendant est montré dans la figure 3.4.b par le hillock 3. Pour le reste de l'hélix, il y a deux hypothèses principales concernant sa formation: *His* a suggéré que les composantes hélicoïdales supérieures et inférieures, y compris l'hélix et l'anthélix, sont dérivées par les hillocks 4 et 5 respectivement, par contre *Streeter* a cru que le 5^{ème} hillock provoque l'hélix tandis que l'anthélix est le produit du hillock 4. Selon la première hypothèse, la région hélicoïdale supérieure, représentée par le hillock 4 dans la figure 3.4.d, semble être sujette aux variations considérables durant la croissance. *Cryptotia* et *Satyr* sont deux anomalies qui montrent le sous-développement de cette région [171]. Parmi les syndromes de ces deux anomalies, nous pouvons voir que l'hélix et l'anthélix de l'oreille sont excessivement grands. D'autre part, l'apparition du *scapha*, la surface concave de la partie libre qui se trouve entre l'hélix et l'anthélix, fournit une marge permettant aux hélix et anthélix d'avoir un certain degré de développement indépendant qui est mieux illustré par l'hypothèse de *Streeter*; suggérant que l'hélix et l'anthélix sont deux composantes séparées.

L'anthélix, mentionné ci-dessus, est sensible aux variations de la région hélicoïdale supérieure, alors que les parties inférieures sont plus constantes. En outre, *Stahl* est la nomination attribuée à une série d'anomalies dans lesquelles le cru supérieur d'anthélix ne se trouve pas dans sa position normale ou il existe un élément additionnel sur le cru. Le cru inférieur d'anthélix avec le cru d'hélix sont, cependant, considérés parmi les moindres composantes variables de l'oreille selon *Streeter*; ce ne sont pas des emplacements courants pour des anomalies.

Dans l'*Otocephaly*, qui est un syndrome accompagné d'une anomalie du pavillon auriculaire, le tragus est absent. Autres anomalies, liées au tragus, peuvent montrer des prolongements ou duplications de tragus [171], indiquant une variation riche dans la forme de cette composante. En revanche, l'antitragus est moins discuté dans l'analyse des anomalies auriculaires.

Le *lobule* est la seule partie de l'oreille qui se compose de graisse plutôt que de cartilage. La forme du lobule peut être changée d'une bonne forme à une forme attachée, et dans quelques oreilles, le lobule est presque inexistant. Ainsi, il peut être, dans certain cas, très discriminant. Cependant, le lobule semble être comme la seule partie de l'oreille qui continue le développement et la déformation pendant le vieillissement de l'être humain [173].

Considération: Comme la forme du lobule se change considérablement avec le vieillissement, en plus, de se trouver fréquemment occlus ou encombré par des boucles, nous ne prêtons pas beaucoup d'attention à cette composante auriculaire.

Considération: Nous trouvons que l'approche la plus logique qui décrit mieux les variations dans la structure de l'oreille réside dans l'arrangement suggéré par *Streeter* (Figure 3.4.b). Bien que les variations observées dans l'hélix et l'anthélix correspondent aux positions des hillocks 4 et 5 respectivement, les arrangements pris par *His* et *Davis* sont aussi très attractifs. Le matériel discuté dans cette section est principalement dérivé de comptes embryologiques et chirurgicaux de l'oreille humaine, revisité comme une nouvelle perspective, pour qu'il soit bien exploité dans l'étape d'implémentation. Nous soutenons particulièrement ces inspirations embryologiques de l'oreille, puisqu'elles relèvent une structure de composante-sage, aident non seulement à concevoir et évaluer des méthodes de reconnaissance, mais renforcent également les prémisses de la validité biométrique de l'oreille. Après avoir montré que l'oreille est, en effet, une structure de composante-sage, nous présentons, dans les prochaines sections, des modèles basés sur des parties spécifiques de l'oreille utilisés dans la reconnaissance automatique.

3.3 Système générique de reconnaissance automatique par oreilles

Un système automatique de reconnaissance biométrique par oreilles peut être vu comme un système typique de reconnaissance des formes, où l'image d'entrée est analysée, prétraitée et réduite à un ensemble de caractéristiques. Cet ensemble de caractéristiques est comparé aux autres ensembles stockés dans la base de données, dont le but est de déterminer l'identité de la personne proclamée. Les étapes saillantes d'un système de reconnaissance classique par oreilles sont illustrées dans le schéma de la figure 3.5 et détaillées dans les sous-sections suivantes.

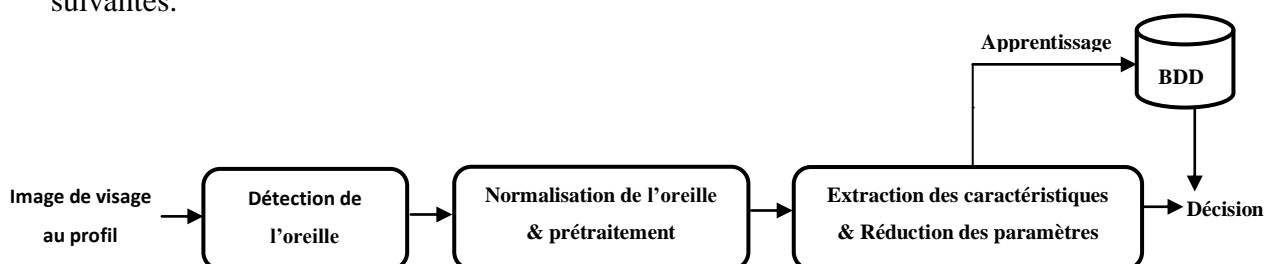


Figure 3. 5 : Différentes positions finales des hillocks, suggérées par différents auteurs [168,169,171].

3.3.1 Détection de l'oreille

La première et plus importante étape, implique la localisation de la position de l'oreille dans une image. Ici, un cadre rectangulaire est typiquement utilisé pour indiquer l'ampleur spatiale de l'oreille dans une image donnée. La détection de l'oreille est une composante critique, puisque les erreurs dans cette étape peuvent miner l'utilité du système biométrique. Plusieurs techniques de détection sont proposées dans le but de localiser l'oreille dans une image donnée. Ces techniques peuvent être catégorisées en quatre approches qui sont [174]:

- a. **Approches basées sur la mise en correspondance:** Dans ce type d'approches, un *Template* d'une oreille typique est construit, ensuite, il est assorti avec tous les endroits de l'image en entrée. L'endroit qui donne le meilleur score de correspondance est considéré comme la région qui contient l'oreille. Le *Template* peut être composé d'une image contenant un contour d'oreille ou d'un ensemble de descripteurs extraits à partir de l'oreille, tel que la réponse d'un ensemble de filtres ou l'histogramme d'un objet incurvé dans le cas où une image 3D est utilisée en identification. La détection basée sur la réponse d'un ensemble de filtres présélectionnés, connue par *la méthode de Viola et Jones* [175], est généralement employée pour détecter les visages dans une image.
- b. **Approches basées sur des modèles géométriques:** Les approches basées sur des modèles géométriques prennent certaines caractéristiques de la forme extérieure de l'oreille et cherchent à trouver les régions qui manifestent de telles caractéristiques. La forme de l'hélix, comme exemple, est habituellement elliptique, donc *la transformée de Hough générale*, utilisée pour détecter les ellipses, peut être utilisée afin de localiser l'oreille dans le contour de l'image. Des caractéristiques extraites par l'utilisation des codes *Chain* peuvent être employées pour classifier chaque courbe obtenue à partir d'une image, à être une courbe liée à l'oreille ou non, telle que l'hélix ou l'anthélix. Un code *Chain* mesure typiquement les orientations locales sur la longueur de la courbe.
- c. **Approches basées sur des opérateurs morphologiques:** Comme la structure de l'oreille est habituellement plus complexe par rapport à la structure de la région restante dans une image de visage de profil, des transformations morphologiques telles que *la transformée du Haut-Chapeau* (Top-Hat transformation) essentiellement soustrait une version morphologique lisse de l'image elle-même, ce qui présente des détails très fins.

d. Approches basées sur la géométrie du visage: Dans une image de visage de profil, le nez peut être facilement détecté comme le point qui contient une courbure très élevée, donc il est possible de contraindre la recherche de l'oreille dans un endroit approprié relatif à la position du nez.

La performance de la détection de l'oreille peut être améliorée par l'utilisation de deux étapes de prétraitement. Tout d'abord, la couleur de la peau humaine est segmentée à partir de l'image de profil, ensuite, l'oreille est détectée facilement dans cet espace de recherche réduit.

3.3.2 Détection de l'oreille

Dans cette étape, l'oreille détectée (segmentée) est soumise à une routine de perfectionnement qui améliore la fidélité de l'image (Ajustement de l'oreille dans la même taille et direction). En plus, l'image de l'oreille peut être soumise à certaines corrections géométriques ou photométriques afin de faciliter l'extraction des caractéristiques et la décision. Dans la plupart des cas, une courbe qui adapte étroitement le contour externe de l'oreille est utilisée dans la normalisation comme montré dans la figure 3.6.

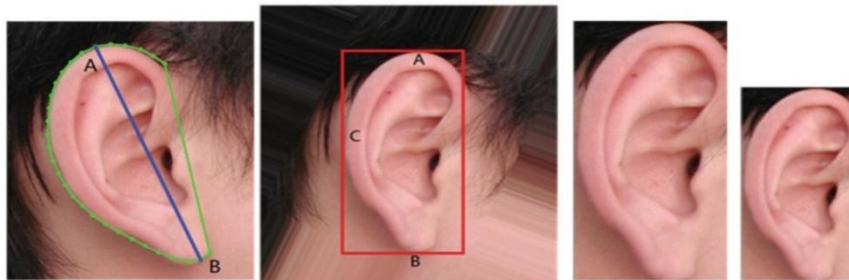


Figure 3. 6 : Exemple de normalisation de l'oreille dans la même direction et taille: (a) Détection du contour et la droite la plus longue (b) Rotation de l'image. (c) et (d) Ajustement de la taille.

3.3.3 Extraction des caractéristiques / Réduction des paramètres

L'oreille segmentée et normalisée peut être directement utilisée dans le module de décision, la plupart des systèmes extraient un ensemble de caractéristiques saillantes pour modéliser l'oreille. L'extraction des caractéristiques se rapporte au processus dans lequel l'oreille normalisée est réduite à un modèle mathématique (ex., un vecteur de caractéristiques) qui récapitule l'information saillante.

Les caractéristiques extraites dans l'étape précédente doivent être comparées avec celles qui sont stockées dans la base de données afin d'établir l'identité de l'oreille en entrée. Dans la forme la plus simple de la correspondance, le système comporte la génération des scores de correspondances en comparant les vecteurs de caractéristiques appartenant à deux images d'oreilles. Le score de correspondance indique la similarité entre deux images d'oreilles. Finalement, les scores de correspondances générés sont utilisés pour se prononcer sur une décision finale. Dans le cas de la vérification, la sortie est de type "oui" ou "non", ce qui indique "genuine" ou "imposteur" respectivement. Par ailleurs en mode identification, la sortie est une liste d'identités classées en fonction de leurs scores de correspondances. Nous présentons dans la section suivante l'historique et les travaux réalisés (état de l'art) dans le domaine de la biométrie par oreilles.

3.4 Approches de reconnaissance biométrique par oreilles

En 1949, *Alfred Iannarelli* s'est intéressé à l'oreille en tant que moyen d'identification des personnes dans le contexte de la criminalistique; il a développé, ensuite, le système d'*Iannarelli* en identification manuelle par oreilles (*The Iannarelli System of Ear Identification*) [176], qui a été employée, pendant plus de 40 ans, par les Agences Légales Américaines. Comme montré dans la figure 3.7, son système consiste essentiellement à prendre un certain nombre de mesures à partir d'un ensemble de repères autour de l'oreille. Dans ce cas, l'image de l'oreille est projetée sur un support standard, où les images sont prudemment alignées en ajustant la position de quelques lignes de références sur le Tragus et le Cru d'hélix. Les images d'oreilles sont alors ajustées dans une taille standard. Les quatre lignes du support de guide séparées par des intervalles de 45°, sont reproduites sur l'image de l'oreille. La figure 3.7 illustre une image d'oreille avec les lignes de guide superposées. L'intersection des lignes de guide et les courbes de l'oreille sont alors marquées. Les mesures d'*Iannarelli* sont calculées en se basant sur les positions de ces marqueurs.

Burge et *Burger* [177,178] étaient les premiers à étudier l'oreille en tant que modalité biométrique dans le domaine de la vision par ordinateurs. Inspirés par les premiers travaux d'*Iannarelli*, ils ont fait une étude théorique et démonstrative sur la viabilité de l'oreille comme une modalité biométrique, en termes d'*unicité* et de *mesurabilité* au cours du temps, et ont implémenté pratiquement, sur ordinateur, un système d'identification à base d'oreilles. Dans ce système, l'objet de chaque oreille est modélisé par un graphe d'adjacences construit à

partir du diagramme de *Vornoi* de la courbe extraite par les segments du contour (les contours sont détectés par le filtre de *Canny*). Pour authentifier les personnes, les auteurs ont introduit un nouvel algorithme de correspondance par graphes, qui prend en compte les segments de courbes incorrectes pouvant être produites dans l'image détectée par contours, en raison de quelques changements, tels que: l'éclairage, l'ombre ou l'occlusion. Ils ont trouvé que ces caractéristiques sont robustes et peuvent être extraites à distance. La figure 3.8 présente les courbes extraites, le diagramme de *Vornoi* et le graphe d'adjacences d'une oreille. Ils ont également identifié l'occlusion par cheveux comme un obstacle majeur dans l'identification par oreilles et ont proposé l'utilisation de l'imagerie thermique pour résoudre ce problème.



Figure 3.7 : Système d'*Iannarelli* en identification manuelle par oreilles.

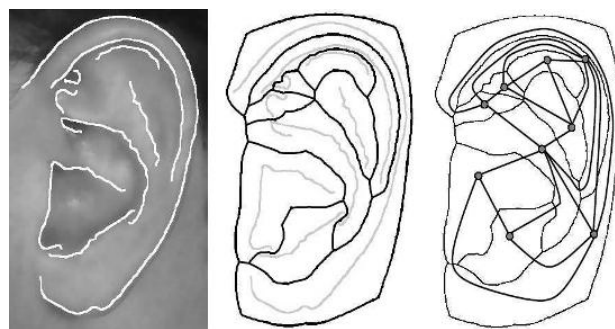


Figure 3.8 Étapes utilisées pour créer le modèle d'oreille basé sur le graphe d'adjacences.

Aujourd'hui, la plupart des systèmes de reconnaissance biométriques basés sur des images d'oreilles 2D, consistent à extraire un ensemble de caractéristiques et de comparer le vecteur extrait avec les modèles enrôlés. A cette base, nous avons divisé les approches de reconnaissance par oreilles en quatre (04) sous classes différentes qui sont: approches basées sur les transformations, approches locales, approches hybrides et approches statistiques (holistiques).

3.4.1 Approches basées sur les transformations

Ces approches traitent l'image entière comme une entité unique et construisent, à l'aide des transformations mathématiques, des caractéristiques correspondantes en se basant directement sur les coordonnées originales de l'image; ce type d'approches présente une grande efficacité dans les conditions contrôlées tandis que leur performance se dégrade considérablement dans la présence des variations de pose, taille ou bruit.

L'approche qui a gagné beaucoup de particularité est la transformée du champ de force (Force Field Transform) proposée par *Hurley et al.* (2002) [179]. Cette transformée suppose que les pixels ont une attraction mutuelle proportionnelle à leurs intensités et inversée par rapport aux distances entre eux, comme la loi de la gravité universelle de *Newton*. Le champ d'énergie associé prend la forme d'une surface lisse avec un certain nombre de pics jointés par des arêtes (voir figure 3.9). Un ensemble de lignes indiquant le gradient de cette force est extrait et utilisé comme un vecteur de caractéristiques. Avec ce modèle, les auteurs ont achevé une performance de 99% au rang-1 sur la base d'images XM2VTS. A partir de ces résultats, *Abdel-Mottaleb et Zhou* (2006) [180] ont utilisé une représentation 3D du champ de force dans le but d'extraire des points liés aux pics du champ de force 3D. Comme le champ de force est convergé vers le profil extérieur de l'oreille, les pics dans la représentation 3D, typiquement, tracent un contour spécifique autour de l'oreille. Par conséquent, la méthode du champ de force est plus robuste contre le bruit par rapport aux méthodes traditionnelles de détection des contours telles que: les filtres de *Sobel*, *Prewitt* et *Canny*. A l'aide de cette approche, *Abdel-Mottaleb et Zhou* ont achevé un taux de reconnaissance au rang-1 égal à 87.93% sur une base d'images privée contenant 103 images de 29 objets (personnes).

Dong et Mu (2008) [181] ont ajouté l'invariance de pose aux contours extraits par la transformée du champ de force. Cela est achevé avec l'analyse NKFDA [182,183] (Null space Kernel Fisher Discriminant Analysis) qui possède la propriété de représenter des relations non-linéaires entre deux nuages de données. Les auteurs ont conduit leurs expérimentations avec la base d'images USTB-4. Avant l'extraction des caractéristiques, la région qui contient l'oreille est découpée en dehors de l'image, tandis que la pose est normalisée manuellement. Pour des variations de pose à 30°, les auteurs ont rapporté un taux de reconnaissance au rang-1 égal à 72.2% et pour des variations de pose à 45°, le taux a chuté à 48.1%.



Figure 3. 9 : Exemple d'extraction des lignes du champ de force à partir d'une image d'oreille [179].

Récemment, *Kumar et Wu* (2012) [184] ont présenté une approche de reconnaissance par oreilles, qui utilise l'information de phase afin de coder les structures locales de l'oreille, à l'aide des filtres de *log-Gabor*. L'information de phase codée est stockée dans une image au niveau de gris. Selon leurs expérimentations, l'approche basée sur les filtres de *log-Gabor* présente des résultats meilleurs par rapport à la transformée du champ de force et aux approches basées sur les points de repères. Cependant, plusieurs combinaisons des filtres de *log-Gabor* sont comparées entre elles. Les performances au rang-k, balance entre 92.06% et 95.93% sur la base d'images IIT Delhi-2.

La structure riche de la morphologie extérieure de l'oreille présente des informations de textures très spécifiques, qui peuvent être mesurées par l'utilisation des filtres de *Gabor*. *Wang et Yuan* (2010) [185] ont extrait des caractéristiques de fréquence locale par l'utilisation de ces filtres. Ensuite, ils ont sélectionné les caractéristiques les plus distinctives en utilisant une analyse discriminante générale. Dans leurs expérimentations sur la base d'images USTB-2, ils ont comparé l'impact de la performance sur différents paramètres du filtre de *Gabor*. Plusieurs combinaisons en orientation et en échelle du filtre sont testées et comparées. Les auteurs ont constaté que l'orientation et l'échelle du filtre n'ont aucun impact important sur la performance. Le taux d'identification total au rang-1 achevé par la méthode de *Wang et Yuan* est de 99.1%. Dans une approche similaire, *Arbab-Zavar et Nixon* (2008) [186] ont mesuré la performance des filtres de *Gabor* sur la base d'images XM2VTS où ils ont rapporté un taux de reconnaissance au rang-1 égal à 91.5%. Les auteurs ont montré que les filtres de *Gabor* présentent certaines faiblesses et limites; en effet, les informations dans les vecteurs de caractéristiques seront perdues par l'occlusion ou autres facteurs disruptifs. Pour résoudre ce

problème, une méthode de classification robuste a été proposée, ce qui a eu pour conséquence d'améliorer le taux d'identification jusqu'à 97.4%.

Abate et al. (2006) [187] ont utilisé un descripteur de *Fourrier* générique pour représenter les caractéristiques invariantes aux rotations et échelles. L'image d'entrée est projetée dans un système à coordonnées polaires et, ensuite, transformée dans un espace fréquentiel. Afin de s'assurer que le centre du système à coordonnées polaires est toujours dans la même position, les images d'oreilles doivent être alignées pour qu'elles puissent être transformées dans un système à coordonnées polaires. La conque agit toujours comme un point référentiel dans l'étape d'alignement, sachant que le point central du système à coordonnées polaires est toujours situé dans la région de la conque. Le système a été évalué sur une base d'images privée qui contient 282 images d'oreilles. Ces images sont prises en deux jours différents avec plusieurs variations de pose. La performance au rang-1 du descripteur de *Fourrier* varie dépendamment avec la pose de chaque angle. Pour l'angle 0° , la performance achevée est de 96%, par contre la performance a été diminuée à 44% avec 15° et à 19% avec 30° .

Dans le travail de *Foopratesiri et Kurutach* (2011) [188], les auteurs ont exploité le concept de la *multi-résolution* appliqué à la transformée de *trace* [189]. Les images d'entrées de la base d'images CMU PIE sont arrangées en série par l'usage de la transformée de *trace* et sont stockées dans des vecteurs de caractéristiques. L'avantage de la transformée de *trace* est que le vecteur de caractéristiques résultant est non invariant aux rotations et aux échelles. En outre, les auteurs ont montré que leur descripteur est aussi robuste contre les variations de pose. Ils ont rapporté totalement un taux de reconnaissance au rang-1 égal à 97%.

Sana et al. (2007) [190] ont utilisé un ensemble de coefficients d'*ondelettes* bien sélectionnés extraits durant la compression par les *Ondelettes de Haar*, afin de représenter les caractéristiques de l'oreille. La transformée d'*ondelette* est appliquée itérativement sur l'image et à quatre niveaux différents. Pour chaque itération, les coefficients dérivés sont stockés dans un vecteur de caractéristiques. Le taux de reconnaissance rapporté par cet algorithme est de 96%, achevé sur les deux bases d'images IITK et Sugor (350 objets).

Basit et Shoaib (2014) [191] ont utilisé la transformée de *curvelete* non-linéaire (Nonlinear Curvelet Transform [192]) avec un classificateur de type k-NN en identification. Les caractéristiques de l'oreille sont extraites par l'application de la transformée de *curvelette* discrète et rapide (FDCT: Fast Discrete Curvelet Transform) sur l'image entière. Le vecteur

de caractéristiques de chaque image se compose de coefficients approximatifs de *curvelettes* et de coefficients secondaires bruts, à huit (08) angles différents. Les auteurs ont achevé des performances au rang-1 à 97.77% et 97.34% sur les deux bases d'images IIT Delhi-1 et IIT Delhi-2 respectivement.

Un système d'extraction des caractéristiques nommé PIFS a été proposé par *De Marsico et al.* (2010) [193]. PIFS, mesure l'inter-similarité dans une image par le calcul d'une transformation affine entre les sous régions similaires d'une image. Dans le but de rendre le système robuste contre les occlusions, les auteurs ont divisé l'image en plusieurs blocks larges et de la même taille. Si un block est occlus, les autres blocks contiennent toujours un ensemble suffisant de caractéristiques distinctives. *De Marsico et al.* ont montré que leur approche est plus robuste par rapport aux autres méthodes d'extraction des caractéristiques sous la présence d'occlusions. Les auteurs ont conduit plusieurs expérimentations dans le but de tester le système dans différents scénarios d'occlusions. Les tests sont effectués sur la base d'images UND (collection E) et les cent (100) premières images de la base FERET. Dans la présence des occlusions, le taux de reconnaissance achevé au rang-1 est égal à 61% (40% est le meilleur taux pour les autres méthodes comparées), tandis que le taux achevé sans occlusion est de 93%.

Les moments invariants sont des mesures statistiques utilisés pour la description des propriétés spécifiques d'un objet. *Wang et al.* (2010) [194] ont composé six (06) vecteurs de caractéristiques différents par l'utilisation de sept (07) moments invariants. Ils ont aussi montré que chaque moment invariant est robuste contre les changements de taille et rotation. Les vecteurs de caractéristiques sont utilisés comme des entrées pour un réseau de neurones de type: *retro-propagation du gradient*, ce dernier a été entraîné pour classifier les vecteurs de caractéristiques créés par des moments invariants. Basés sur une base d'images privée contenant 60 images, ils ont rapporté un taux de reconnaissance au rang-1 égal à 91.8%. *Wang et Yuan* (2009) [195] ont comparé les particularités de différentes méthodes d'extraction des caractéristiques sur la base d'images USTB-1. Ils ont comparé les performances au rang-1 des descripteurs de *Fourrier*, la transformée de *Gabor*, les moments invariants et les caractéristiques statistiques. Ils ont conclu que le meilleur taux d'identification (égal à 100%) a été achevé par la combinaison des moments invariants avec la transformée de *Gabor*.

3.4.2 Approches locales

Les algorithmes basés sur des correspondances locales cherchent à trouver, dans l'image, des points-clés spécifiques, ensuite les caractéristiques sont extraites autour de ces points spécifiques. Les caractéristiques locales ont montré plus de robustesse dans les conditions non-contraintes par rapports aux modèles holistiques. Malgré le pouvoir de ces algorithmes de gérer le problème de variations de pose et taille, ils ont une difficulté de trouver une méthode précise et exacte qui localise les points-clés.

La transformée de caractéristiques visuelles invariantes à l'échelle (SIFT: Scale Invariant Feature Transform) est considérée comme un moyen robuste en extraction des repères dans les images contenant des petites variations de pose et en différentes conditions de luminosité [196]. Les repères SIFT contiennent une mesure d'orientation locale; qui peut être également utilisée pour l'estimation de la rotation. *Bustard et Nixon* (2010) [197] ont montré que les repères SIFT peuvent gérer les variations de pose jusqu'à 20°. Cependant, il n'est pas trivial d'attribuer des repères SIFT avec leurs contreparties exactes, spécialement dans la présence des variations de pose. Dans les régions de l'image fortement structurées dont la densité et la redondance des repères SIFT sont élevées, une affectation exacte n'est pas possible. Par conséquent, les repères doivent être filtrés avant de commencer une comparaison réelle. *Arbab-Zavar et al.* (2007) [198] ainsi que *Badrinath et Gupta* (2009) [199] ont entraîné un modèle de repères référentiel, qui contient uniquement un nombre restreint de repères non-redondants. Ce modèle référentiel est utilisé pour filtrer les repères SIFT, qui sont initialement détectés dans l'oreille de test et de référence. Ayant les repères filtrés, il est possible d'attribuer chaque repère avec sa contrepartie assortie. La figure 3.10 illustre un exemple de repères SIFT extraits à partir des images d'oreilles, utilisées comme des données d'apprentissage pour le modèle référentiel. Comme *Arbab-Zavar et al.* ont aussi utilisé la base d'images XM2VTS dans l'évaluation, leurs résultats peuvent être directement comparés aux performances rapportées par *Bustard et Nixon*. *Arbab-Zavar et al.* ont achevé une performance au rang-1 égale à 91.5%. Dans une approche plus récente de *Bustard et Nixon*, la performance a été améliorée jusqu'à 96%.

Kisku et al. (2009) [200] ont résolu le problème d'affectation correcte des repères par la décomposition de l'image d'oreille en différents segments de couleurs. Les repères SIFT sont extraits séparément de chaque segment, ce qui réduit la possibilité d'attribuer des repères

SIFT qui ne représentent pas les mêmes caractéristiques. Avec cette approche, les auteurs ont atteint une performance au rang-1 égale à 96.93%.

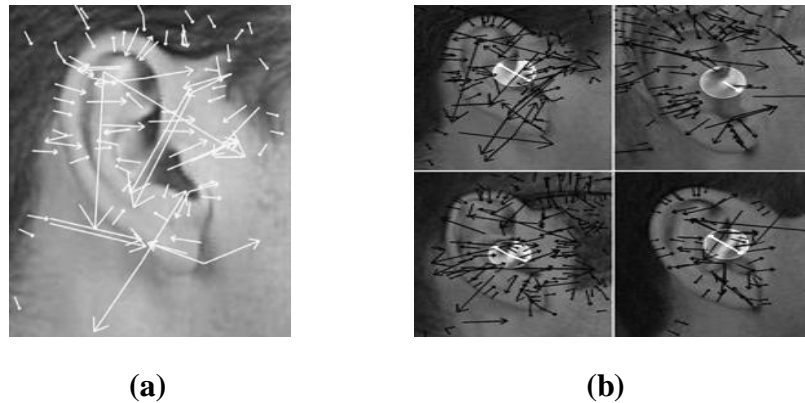


Figure 3. 10 : (a) Repères SIFT d'une image d'oreille (b) Crus d'hélix détecté par SIFT dans quatre images différentes.

Une approche récente proposée par *Prakash et Gupta* (2013) [201] fusionne les caractéristiques robustes et accélérées (SURF: Speeded Up Robust Features) [202] à partir de plusieurs images du même objet. Les auteurs ont proposé d'employer plusieurs images en apprentissage et de stocker tous les points de caractéristiques SURF pour créer un vecteur de caractéristiques concaténé. Cet ensemble de caractéristiques est ensuite utilisé pour entraîner un classificateur de type k-plus proche voisin (k-NN) afin d'assigner deux points de caractéristiques corrélées. Si la distance entre deux points de caractéristiques SURF est inférieure à un seuil donné, elles sont considérées corrélées. L'évaluation de cette approche est faite sur la base d'images UND (collection E) et les deux sous-ensembles de la base d'images IIT Kanpur. *Prakash et Gupta* ont testé l'influence de différents paramètres du descripteur SURF et du classificateur k-NN. En fonction de la composition de ces paramètres, l'EER varie entre 6.72 et 5.25%.

Choras (2008) [203] a proposé un ensemble de méthodes géométriques pour l'extraction des caractéristiques, inspiré par les premiers travaux d'*Iannarelli*. Il a proposé quatre (04) manières différentes de localisation des caractéristiques autour des contours de l'oreille. *La méthode des cercles concentriques* utilise la conque comme point de référence pour un certain nombre de cercles ayant des rayons prédéfinis. L'ensemble des points de l'intersection des cercles avec les contours de l'oreille est utilisé comme un vecteur de caractéristiques (Voir la figure 3.11). Une extension de cette méthode est nommée: *la méthode de traçage des contours* qui consiste à utiliser les bifurcations, les points finaux et les points d'intersection

entre les contours de l'oreille comme des caractéristiques additionnelles. Dans *l'approche de représentation par angle*, *Choras* a tracé des cercles concentriques autour de chaque point central de chaque contour. Ensuite, il a utilisé des angles entre les points centraux et les cercles concentriques. Finalement, les points d'intersection sont utilisés comme des caractéristiques. La dernière méthode, nommée: *la méthode de proportion par triangle*, détermine les distances normalisées entre les points de références et les utilise pour la description de l'oreille. *Choras* a fait des études sur différentes bases d'images où il a rapporté des taux d'identification entre 86.2% et 100% sur une base d'images contenant 12 objets et un taux de faux rejets entre 0 et 9.6% sur une base d'images contenant 102 images d'oreilles.

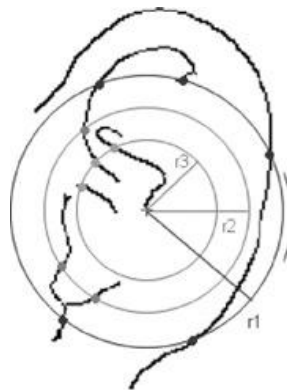


Figure 3. 11 : Représentation symbolique de la méthode des cercles concentriques.

Des approches similaires qui utilisent l'aspect de la proportion entre les points référentiels sur les contours de l'oreille sont proposées par *Mu et al.* (2005) [204] et *Rahman et al.* (2007) [205]. *Mu et al.* ont atteint un taux de reconnaissance au rang-1 égal à 85% avec la base d'images USTB-2. Tandis que *Rahman et al.* ont évalué leur approche sur une base d'images privée contenant 350 images de 100 personnes, ils ont rapporté un taux de reconnaissance au rang-1 de 90%.

Sibai et al. (2013) [206] ont défini un vecteur de caractéristiques géométriques de sept (07) éléments et ont entraîné un réseau de neurones de type *rétro-propagation du gradient*, pour identifier les personnes par leurs oreilles. Le vecteur de caractéristiques se compose de: la longueur de l'hélix, la longueur de l'anthélix, la longueur-1 (Plus longue distance au-dessus du périmètre de l'oreille vers le plus haut point au fond), la longueur-2 (Longueur de la ligne où l'oreille attache le visage), la largeur (Trait horizontal, droit au-dessus de l'auricule), la longueur du tragus et la longueur du lobule, comme montré dans la figure 3.12. Les auteurs

ont entraîné et testé le réseau de neurones avec une base d'images privée qui contenant 51 images d'oreilles de différentes personnes. Les expériences de la simulation sont faites avec différents réseaux de neurones; plusieurs couches, plusieurs neurones par couche, plusieurs fonctions d'activation, avec et sans bruit. Le meilleur taux d'identification achevé est de 95% obtenu avec 03 couches, 18 neurones et en présence de quelques bruits.

Le motif binaire local (LBP: Local Binary Pattern) est une technique d'extraction des caractéristiques au niveau pixel. LBP codifie le voisinage local d'un pixel en stockant la différence entre le pixel examiné et ses voisins. *Guo et al.* (2008) [207] ont extrait les codes LBP à partir des images d'oreilles brutes et ont créé des histogrammes décrivant la distribution locale des codes LBP. Ensuite, un réseau de neurones cellulaire entraîné, a été utilisé pour distinguer entre les codes LBP de différents objets de la base d'images USTB-2.

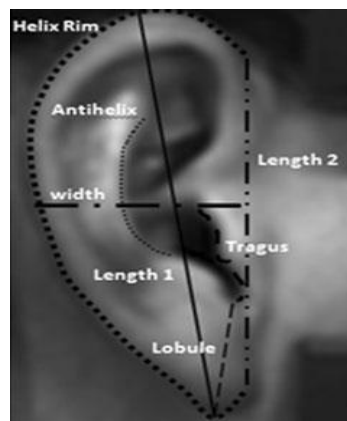


Figure 3. 12 : Caractéristiques géométriques sélectionnées par la méthode de *Sibai et al.* (2013) [206].

Wang et Yan (2011) [208] ont utilisé l'analyse discriminante linéaire (LDA) afin de réduire la dimensionnalité des vecteurs contenant les codes LBP. Puis, ils ont quantifié la similarité entre les vecteurs des caractéristiques à l'aide de la distance *Euclidienne*. *Wang et Yan* ont évalué leur approche sur la base d'images USTB-2 et ont rapporté une performance au rang-1 égale à 100%.

3.4.3 Approches hybrides

Les approches hybrides permettent d'associer les avantages des méthodes basées sur les transformations et les méthodes locales en combinant la détection des caractéristiques structurales avec l'extraction des caractéristiques d'apparences locales. Elles permettent

d'augmenter la stabilité de la performance de la reconnaissance lors du changement de la pose, la taille et l'éclairage.

L'approche proposée par *Jeges et Mate* (2007) [209] a deux étapes dans l'extraction des caractéristiques. Premièrement, ils ont généré un modèle de contour moyen à partir d'un ensemble d'images d'apprentissage. Ces contours représentent le contour extérieur de l'hélix ainsi que les contours de l'anthélix, le cru d'hélix et la conque. Ensuite, chaque image est enregistrée en déformant le modèle de l'oreille jusqu'à ce qu'il adapte les contours actuels affichés dans l'image de l'oreille utilisée en test. Les paramètres de déformation, qui sont nécessaires pour la transformation, représentent la première partie du vecteur de caractéristiques. Ce dernier est accompli par l'ajout d'un ensemble de points de caractéristiques additionnelles liés à l'intersection entre un ensemble d'axes prédéfinis et les contours principaux transformés. Les axes représentent le contour unique de l'oreille. La figure 3.13 montre une image d'oreille améliorée par les bords avec des contours ajustés ainsi que les axes additionnels pour l'extraction des points référentiels. Les auteurs ont achevé un EER de 5.6% sur une base contenant des images coupées et sans aucune variation de pose.

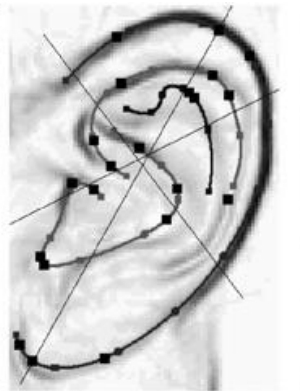


Figure 3.13 : Exemple d'identification par la méthode de *Jeges et Mate* (2007) [209].

Liu et Yan (2007) [210] ont combiné la vue de face et la vue arrière de l'oreille en extraction des caractéristiques par l'utilisation de la méthode du rapport de triangle et les descripteurs des moments de *Tchebichef*. Les moments de *Tchebichef* sont des ensembles de fonctions de moments orthogonaux basés sur des polynômes discrets de *Tchebichef* qui sont considérés comme des descripteurs de caractéristiques [211]. La vue arrière de l'oreille est décrite par un certain nombre de lignes perpendiculaires au plus long axe dans le contour de l'oreille. Ces lignes mesurent le diamètre local du pavillon auriculaire sur des points

prédéfinis. La performance au rang-1 de cette approche hybride atteint 97.5%. Si uniquement la vue de face est utilisée, la performance atteint 95%. Par contre, si uniquement la vue arrière est utilisée, *Liu et Yan* ont atteint une performance de 86.3%.

Lu et al. (2006) [212] ainsi que *Yuan et Mu* (2007) [213] ont utilisé le contour actif pour extraire le bord de l'oreille. Les premiers ont travaillé avec des images d'oreilles coupées manuellement de 56 personnes en différentes poses. Un extracteur de caractéristiques enregistre un ensemble de points sélectionnés sur le contour de l'oreille ainsi que leurs distances par rapport au tragus. Avant d'appliquer un classificateur linéaire, la dimensionnalité du vecteur de caractéristiques est réduite par l'usage de l'analyse en composante principale (PCA). *Lu et al.* ont comparé la performance au rang-1 seulement si l'oreille gauche ou droite est utilisée dans l'identification. Ils ont également montré que lors de l'utilisation des deux oreilles, la performance a été améliorée et passée de 93.3% à 95.1%. Sur la base d'images USTB-3, *Yuan et Mu* ont rapporté une performance au rang-1 de 90% où la rotation principale est inférieure à 15°. Pour des angles de rotation entre 20° et 60°, la performance a été diminuée jusqu'à 80%.

3.4.4 Approches statistiques / holistiques

Les approches statistiques sont basées sur l'information des pixels; tous les pixels de l'image sont traités comme un vecteur unique. En effet, le nombre total de ces pixels représente la taille du vecteur. La plupart des méthodes de cette approche utilisent un autre espace de représentation (sous-espace) pour réduire le nombre de pixels et éliminer les redondances. L'analyse en composante principale (PCA), l'analyse discriminante linéaire (LDA) et l'analyse en composante indépendante (ICA) sont les méthodes les plus populaires utilisées pour réduire les dimensions et sélectionner les bons paramètres.

Victor et al. (2002) [214] étaient le premier groupe de recherche qui a transféré l'idée qui consiste à utiliser les sous-espaces (Eigen-spaces) de la reconnaissance par visages vers la reconnaissance par oreilles. Ils ont suggéré que la performance de l'oreille, en tant que modalité biométrique, est moins inférieure que la performance du visage. Cela peut être dû au fait que dans leurs expérimentations, ils ont considéré que les oreilles gauche et droite sont symétriques. Ils ont utilisé une oreille en apprentissage et l'autre oreille en test ce qui permet, dans ce cas, de réduire la performance de la PCA; la performance rapportée au rang-1 est de 40%. Avec une performance au rang-1 de 72% sur la base d'images UND (collection E),

Chang et al. (2003) [215], ont rapporté une performance significativement meilleure que *Victor et al.* *Alaraj et al.* (2010) [216] qui ont publié une autre étude, où la PCA a été utilisée pour représenter les caractéristiques en reconnaissance par oreilles. Dans leur approche, un réseau de neurones multicouches de type *retro-propagation du gradient* a été entraîné pour classifier les vecteurs de caractéristiques à base de PCA. La performance observée au rang-1 est de 96%, et par conséquent les résultats ont été améliorés par rapport à ceux de *Victor et al.* et *Chang et al.* Cependant, il faut noter que ce résultat est uniquement basé sur un sous ensemble d'une collection de la base d'images UND, qui contient 85 images de 17 objets.

Zhang et Mu (2008) [217] ont mené des études sur l'efficacité des méthodes statistiques en combinaison avec les classificateurs. Ils ont montré, sur la base d'images USTB-1, que l'analyse en composante indépendante (ICA) est plus efficace par rapport à la PCA. Les auteurs ont utilisé, premièrement, les deux méthodes PCA et ICA pour réduire la dimensionnalité des images en entrée et, ensuite, ont entraîné un classificateur de type SVM pour grouper les vecteurs de caractéristiques. En outre, l'influence de différentes tailles du vecteur de caractéristiques est mesurée. Selon les différentes tailles testées, la performance au rang-1 avec PCA varie entre 85% et 94.12%, tandis que la performance avec ICA varie entre 91.6% et 100%.

Xie et Mu (2008) [218] ont proposé un algorithme amélioré, nommé: le prolongement localement linéaire (LLE: Locally Linear Embedding), pour réduire la dimensionnalité des caractéristiques de l'oreille. LLE est une technique permettant de projeter un ensemble de points dans un espace de haute dimension vers un système de coordonnées à faible dimension, tout en préservant les relations spatio-temporelles entre les points de données individuelles. Cela nécessite que certains points de données doivent être marqués d'une certaine manière, de sorte que leur rapport reste toujours fixe. La version améliorée du LLE proposée par *Xie et Mu*, élimine ce problème par l'utilisation d'une fonction de distance différente. En outre, *Xie et Mu* ont montré que l'algorithme LLE est plus performant que la PCA, si les données d'entrée contiennent des variations de pose. Leurs études sont faites sur la base d'images USTB-3, ils ont montré que la performance au rang-1 du LLE régulier est de 43%. Cette performance a été améliorée par leur méthode jusqu'à 60.75%. En absence de variation de pose, l'approche LLE améliorée atteint une performance de 90%.

Dans l'approche proposée par *Nanni et Lumini* (2007) [219], les auteurs ont utilisé la sélection séquentielle croissante et flottante (SFFS: Sequential Forward Floating Selection), qui est une méthode itérative et statistique pour la sélection des paramètres, bien connue dans le domaine de la reconnaissance des formes. SFFS cherche à trouver le meilleur ensemble de classificateurs par la création d'un ensemble de règles, qui garde la meilleure forme de l'ensemble courant de caractéristiques. Les ensembles sont créés par l'ajout d'un classificateur à chaque itération et l'évaluation de sa puissance discriminative se fait en utilisant une fonction de fitness prédéfinie. Si le nouvel ensemble de règles surpasse la version précédente, la nouvelle règle est ajoutée au dernier ensemble. Leurs expériences ont été effectuées sur la collection E de la base d'images UND, et les classificateurs individuels ont été fusionnés en utilisant la règle de la somme pondérée. SFFS sélectionne les sous-fenêtres les plus discriminantes qui correspondent à l'ensemble de règles le plus convenable. *Nanni et Lumini* ont rapporté un taux de reconnaissance au rang-1 de 80% et un taux de 93% au rang-5. L'EER varie entre 6.07% et 4.05% selon le nombre de sous-fenêtres utilisé.

Yiuzono et al. (2002) [220] ont considéré que trouver les caractéristiques saillantes dans une image d'oreille est un problème d'optimisation; ils ont appliqué une recherche locale à l'aide d'un algorithme génétique pour résoudre le problème itérativement. Ces auteurs ont sélectionné des sous-fenêtres locales avec des tailles variantes considérées comme des paramètres initiaux de l'algorithme génétique. Ils ont présenté des résultats élaborés, qui décrivent le comportement d'une recherche génétique locale sous différents paramètres, tels que différentes méthodes de sélection et différents nombres de chromosomes. Sur une base d'images privée de 110 objets, ils ont rapporté un taux de reconnaissance au rang-1 égal à 100%.

Yaqubi et al. (2008) [221] ont utilisé des caractéristiques obtenues par une combinaison entre la position et les détecteurs de contours à balance tolérante, sur plusieurs positions et orientations de l'image. Cette méthode d'extraction des caractéristiques, nommée: modèle HMAX, est inspirée par le cortex visuel des primats qui combine des caractéristiques simples à des entités sémantiques plus complexes. Les caractéristiques extraites sont classifiées avec un SVM et un k-NN. La performance au rang-1 sur une petite base d'images contenant 180 images d'oreilles de six (06) personnes variée entre 62% et 100% selon les paramètres utilisés.

Moreno et al. (1999) [222] ont implémenté un extracteur de caractéristiques qui localise sept (07) repères autour de l'image de l'oreille, ces repères correspondent aux points saillants du système d'*Iannarelli*. En plus, les auteurs ont obtenu un vecteur de morphologie qui présente l'oreille comme un ensemble de points. Ces deux vecteurs de caractéristiques sont utilisés comme des entrées pour un classificateur de multiples réseaux de neurones. Ils ont comparé la performance de cette technique avec différentes méthodes de fusion. La base d'images propriétaire utilisée en test contient 168 images de 28 objets manuellement découpées. Le meilleur taux de reconnaissance obtenu au rang-1 est égal à 93%.

Gutierrez et al. (2010) [223] ont divisé l'image de l'oreille en trois parties de même taille. La partie haute représente l'hélix, la partie du milieu représente la conque, et la partie basse modélise le lobule. Chaque partie de ces sous-images est décomposée par la transformée d'ondelettes, puis introduite dans un réseau de neurones modulaire. Dans chaque module du réseau, plusieurs intégrateurs et fonctions d'apprentissage sont utilisés. Les résultats de chaque module ont fusionnés dans la dernière étape dans le but d'obtenir une décision finale. Selon les différentes combinaisons entre l'intégrateur et la fonction d'apprentissage, les résultats varient entre 88.4% et 97.47% au rang-1, sur la base d'images USTB-1. La meilleure performance au rang-1 a été achevée avec la distance de *Sugeno* et le gradient conjugué.

Dans le travail de *Naseem et al.* (2008) [224], les auteurs ont proposé un algorithme de classification général basé sur la théorie de la sensation compressive. Ils ont supposé que la plupart des signaux de la nature sont compressibles et que n'importe quelle fonction de compression aboutit à une représentation clairsemée (Sparse Representation) de ce signal. Dans leurs expérimentations sur les deux bases d'images UND et FEUD, *Nasseem et al.* ont montré que leur méthode basée sur la représentation clairsemée est robuste contre les variations de pose et les différentes conditions d'éclairage. La performance au rang-1 varie entre 89.13% et 97.83% selon la base d'images utilisée en expérimentation.

Zhang et al. (2014) [225] ont proposé une approche d'extraction des caractéristiques, basée sur l'information de l'orientation des ondelettes de *Gabor* et la représentation clairsemée. Le filtre de *Gabor* extrait l'information d'orientation de l'oreille en différentes directions et décrit l'information de son contour. Les caractéristiques visuelles de l'orientation de *Gabor* sont combinées avec les caractéristiques de la représentation clairsemée. L'approche proposée a achevé des performances très intéressantes et une grande efficacité dans les conditions non-

contraintes. Le meilleur taux de reconnaissance achevé est de 99.75% sur la base d'images USTB-4 avec 0° en orientation, tandis que ce taux a été diminué vers 93.67% et 52.41% avec 28° et 45° en orientation respectivement.

3.5 Défis ouverts et futures application

La plupart des publications récentes sur la reconnaissance 2D de l'oreille montrent que le défi principal de cette technologie réside dans l'identification des individus dans les environnements non-contraints. Cela inclut des applications de surveillance intelligente ainsi que l'identification légale des malfaiteurs, sur des séquences de vidéo ou pour les systèmes de contrôle des frontières. Traditionnellement, ces domaines d'application font partie des systèmes d'identification par visages, mais, comme l'oreille se trouve sur le coté du visage, elle peut fournir des informations additionnelles et précieuses pour compléter les images faciales.

Les systèmes d'identification multimodaux combinant le visage et l'oreille peuvent servir comme des moyens pour achever l'invariance de pose et donner plus de robustesse contre l'occlusion dans les environnements non-contraints. Dans la plupart des lieux publics, les caméras de surveillance sont situées au-dessus afin de capturer autant de personnes que possible et de les protéger contre le vandalisme. En plus, la plupart des personnes ne regardent pas directement la caméra, donc dans la plupart des cas aucune image frontale de la personne ne sera disponible. Ce facteur pose des problèmes sérieux aux systèmes biométriques utilisant les traits du visage en identification; si le visage n'est pas visible à partir d'un angle frontal, l'oreille peut fournir de précieuses caractéristiques supplémentaires dans ces scénarios.

En raison de la proximité physique du visage et de l'oreille, il existe également de nombreuses possibilités pour fusionner ces deux modalités biométriques. Les images du visage et de l'oreille peuvent être fusionnées au niveau capteur, niveau caractéristique ou niveau décision. Dans la perspective de cette application, il reste quelques défis non résolus, pour les futures recherches dans ce domaine.

3.5.1 Localisation automatique de l'oreille

Le fait que de nombreux systèmes présentés dans la littérature utilisent des images d'oreilles pré-segmentées, cela prouve que la détection automatique des oreilles, spécialement dans les images réelles, reste toujours un problème non-résolu. Si les systèmes de

reconnaissance par oreilles doivent être implémentés dans des systèmes d'identification automatique alors une approche de détection rapide et efficace doit être utilisée. Comme une première étape vers cet objectif, quelques groupes de recherches ont publié des collections de données, qui simulent les variations typiques dans des environnements non-contrôlés tels que: différentes conditions d'éclairage, pose et occlusion. En se basant sur cet ensemble de collections, les approches existantes et les futures approches de détection des oreilles doivent être testées dans des conditions réelles dans le but d'améliorer leur fiabilité.

3.5.2 Occlusion et les variations de la pose

Contrairement au visage, l'oreille peut être partiellement ou complètement couverte par des cheveux ou par autres éléments, tels que: les coiffes artificiels, les prothèses auditives, les bijoux ou les écouteurs. En raison de la surface convexe de la partie extérieure de l'oreille, des parties de celle-ci peuvent être aussi occluses s'il y a des changements de pose. Dans certaines publications, la robustesse contre l'occlusion est explicitement abordée, mais il n'y a aucune étude sur l'effet de certains types d'occlusion comme les cheveux ou les boucles sur le taux de reconnaissance de l'application biométrique. Autrefois, la disponibilité des bases de données publiques contenant des images d'oreilles occluses est susceptible à stimuler le développement des solutions pour des algorithmes robustes et invariables contre les variations de pose, dans la détection et extraction des caractéristiques de l'oreille.

Considération: Selon notre connaissance, il n'y a aucune étude sur la visibilité externe de l'oreille en différents environnements publics. Afin de développer des algorithmes de détection et d'identification d'oreilles, plus d'informations et détails sur les parties de l'oreille communément occluses sont nécessaires.

L'occlusion liée à la variation de pose est un autre défi à relever dans les systèmes de reconnaissance par oreilles. Similaire à la reconnaissance par visages; des parties de l'oreille peuvent devenir occluses si la pose est changée. Récemment, quelques méthodes d'extraction des caractéristiques sont proposées, elles sont robustes contre les variations de pose à certains degrés d'inclinaison. Cependant, le problème n'est pas encore résolu complètement. Une autre possibilité pour compenser le problème de la variation de pose réside dans l'utilisation des modèles 3D au lieu d'images photographiques en profondeur, quoique l'utilisation du modèle 3D ne soit pas préférable pour les applications en temps-réel et nécessite souvent des équipements spécifiques.

3.5.3 Évolutivité des systèmes de reconnaissance par oreilles

Actuellement, les bases de données uniquement disponibles se constituent de moins de 10 000 images d'oreilles. Dans les environnements réalistes, la taille de la base de données est significativement grande, ce qui rend la recherche exhaustive dans les scénarios d'identification non faisables. Par conséquent, non seulement la performance mais aussi la vitesse de la comparaison dans les systèmes d'identification par oreilles sera intéressante dans les futures recherches.

Afin de rendre l'identification par oreilles applicable pour les systèmes à grande échelle, des recherches exhaustives doivent être remplacées par des structures de données appropriées permettant une complexité logarithmique durant la recherche. Cela peut, par exemple, être réalisé en explorant la possibilité d'organiser les modèles de l'oreille dans des structures d'arbres de recherche.

3.5.4 Comprendre la symétrie et le vieillissement

La reconnaissance par oreilles est considérée comme l'un des nouveaux champs de recherches dans le domaine de la biométrie, mais la symétrie des oreilles gauche et droite n'a pas encore été complètement comprit. Les études d'*Iannarelli* indiquent que certaines caractéristiques de la partie extérieure de l'oreille peuvent être héritées et le vieillissement affecte légèrement l'apparence extérieure de l'oreille. Les deux hypothèses peuvent être confirmées par des études plus récentes, mais en raison d'un manque de données, l'effet d'héritage et du vieillissement sur l'apparence extérieure de l'oreille n'a pas encore été complètement compris. En outre, il n'y a aucune étude à grande échelle sur la relation de la symétrie entre les deux oreilles: gauche et droite.

Par conséquent, un autre champ intéressant pour des futures recherches peut gagner une compréhension plus profonde sur l'effet d'héritage de n'importe quelle symétrie sur les particularités du modèle biométrique. D'ailleurs, des études à long-terme sur l'effet du temps sur les modèles de l'oreille sont nécessaires pour obtenir une meilleure compréhension de la performance de cette modalité biométrique.

3.6 Conclusion

Dans ce chapitre, nous avons présenté une étude détaillée sur l'état de l'art de la reconnaissance biométrique à base d'oreilles, couvrant des techniques de détection et de reconnaissance. La reconnaissance par oreilles est un nouveau domaine de recherche. Bien qu'il existe un certain nombre d'approches prometteuses, aucune d'entre elles n'a été évaluée avec des scénarios réalistes, qui comprennent des facteurs perturbateurs tels que: les variations de pose, l'occlusion et les différentes conditions d'éclairage. Dans les approches récentes, ces facteurs sont pris en considération, mais beaucoup de recherches sur ceux-ci sont exigées jusqu'à ce que les systèmes de reconnaissance par oreilles puissent être utilisés dans la pratique. La disponibilité de quelques bases de données d'essai appropriées, collectées dans des scénarios réalistes, contribuera plus loin à l'évolution de l'oreille comme une caractéristique biométrique.

Au cours de cette recherche, nous avons collecté un aperçu structuré sur les approches existantes dans la détection et la reconnaissance, ainsi que les problèmes non-résolus de la reconnaissance par oreilles dans le contexte de la surveillance intelligente et la criminalistique, que nous considérons comme les applications les plus importantes en biométrie de l'oreille. Nous pensons que cette nouvelle modalité est une extension précieuse pour les systèmes de reconnaissance par visages, sur le chemin de l'identification automatique sous la présence des variations de pose.

Chapitre 4

Caractérisation Biométrique par Descripteurs de Texture Locaux

4.1 Introduction

Après avoir introduit, dans les deux chapitres précédents, les principales méthodes de caractérisation (extraction des caractéristiques) en lien avec le visage et l'oreille, nous présentons dans ce chapitre les techniques que nous avons utilisées dans le but d'extraire des informations biométriques texturées.

L'analyse de texture réfère à la discipline de l'analyse d'images qui s'intéresse à la description des caractéristiques de l'image par des attributs texturaux. Cependant, il n'existe pas une définition universellement acceptée de ce qui est une texture de l'image, en général, plusieurs chercheurs utilisent des définitions différentes selon leurs domaines d'intérêt [226]. Dans ce chapitre, la texture est considérée comme la variation spatiale d'intensités de pixels, ce qui est une définition largement utilisée et acceptée dans le domaine d'imagerie. L'objectif de ce chapitre est de présenter une revue bibliographique des méthodes d'analyse de texture existantes, avec un intérêt pour les techniques utilisées en biométrie.

Nous nous sommes intéressés dans cette étude par les descripteurs de texture locaux inspirés principalement par la technique des *motifs binaires locaux* (LBP pour Local Binary Patterns).

4.2 Motivation

Le choix de descripteurs de texture locaux est motivé par: (i) les études en neuropsychologie qui ont bien montré que le système neurologique humain utilise les deux informations: holistique (ex., la forme globale de l'objet) et locale (ex., la texture ou les points de repères) pour différencier et reconnaître les individus (voir chapitre 2, section 2.3. (ii) les développements récents qui ont montré que les motifs binaires locaux et ses variantes fournissent des résultats *exceptionnels* dans la représentation et l'analyse des caractéristiques; ils ont la capacité à dériver facilement des modèles de caractéristiques efficaces qui

combinent dans un vecteur de données unique la forme globale de l'objet analysé et la texture locale de son apparence. Les représentations obtenues, à l'aide de ces descripteurs, rapportent des performances très excellentes. Donc, (iii) l'utilisation de ce type de descripteurs dans l'analyse, la modélisation, et la reconnaissance biométrique est considérée comme un *excellent* choix, puisqu'ils imitent les capacités de l'être humain en reconnaissance de visages ou objets, par ses particularités en codification des composantes à basses et hautes fréquences qui contribuent à la description locale et globale respectivement.

Avant d'entrer dans les détails de ces descripteurs et leur fonctionnement, nous présentons, dans la section suivante, quelques notions fondamentales sur l'analyse de texture.

4.3 Notions fondamentales sur l'analyse de texture

L'analyse de texture est un domaine actif de recherches intenses et de vaste littérature. Néanmoins, il existe deux sujets fondamentaux qui ne sont pas résolus jusqu'à présent: (i) la définition de la notion de texture; et (ii) la création d'une taxonomie significative et sans ambiguïtés des descripteurs de texture existants. Le manque de solutions satisfaisantes à ces deux sujets clés est un point de faiblesse sérieux qui limite le progrès de la discipline. Nous croyons fortement que des avancées significatives dans le domaine peuvent être réalisées. Dans les paragraphes qui suivent, nous discutons des deux sujets en détails.

4.3.1 Définition de la texture

La texture est un terme largement utilisé dans la vision par ordinateurs, et c'est plutôt surprenant qu'un tel concept omniprésent n'a pas trouvé un consensus général sur une définition explicite. *Ahonen et al.* (2009) [227] ont correctement noté que c'est peut-être une des raisons pour lesquelles ni une théorie unificatrice, ni un cadre de descripteurs de texture ont été proposés jusqu'à présent.

La racine du mot latin (*texere* = à tisser) suggère que la texture est un peu liée à l'interaction, la combinaison, et l'imbrication d'éléments dans un ensemble complexe (Figure 4.1). Le concept de texture comme la propriété visuelle d'une surface, cependant, est assez subjectif et imprécis. Nous pouvons reconnaître la texture quand nous la voyons, mais la définition d'une manière formelle est beaucoup plus difficile. Certainement il y a quelques attributs de texture qui sont largement convenus: cette texture est la propriété d'une zone (et pas d'un point), elle est liée à la variation de l'apparence, elle dépend fortement à l'échelle

d'une image, et elle est perçue comme la combinaison de certains modèles basiques. Par exemple, *Davies* (2008) [228] affirme que la plupart des gens considèrent la texture comme « un motif aléatoire et régulier en même temps ». *Petrou et Sevilla* (2006) [229] considèrent la texture comme « une variation de données à des échelles plus petites que l'échelle d'intérêt ». En fait, beaucoup de définitions sont proposées dans la littérature: le lecteur peut trouver un petit recueil dans la référence [226] ; malheureusement, aucune de ces définitions n'a suscité un consensus général, principalement parce qu'il n'y a pas de modèle mathématique formel sur lequel nous pouvons en déduire une définition générale quantitative.

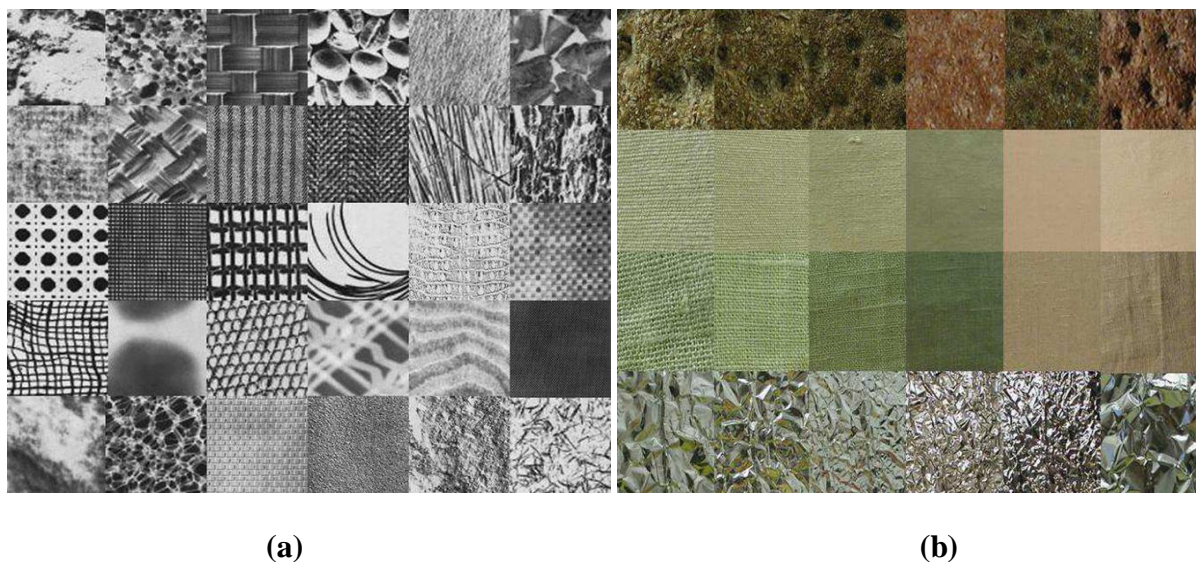


Figure 4. 1 : Exemples de textures: (a) base d'images *Brodatz* (b) base d'images *KTH-TIPS2*.

Considération: Dans le cadre de cette thèse, nous avons considéré la texture comme la variation spatiale d'intensités de pixels; c'est une définition largement utilisée et acceptée spécialement dans le domaine de l'imagerie.

4.3.2 Catégorisation des descripteurs de texture

Le deuxième sujet critique qui est, en fait, une conséquence directe du premier, il concerne le développement d'une taxonomie pour les descripteurs de texture. Plusieurs tentatives pour classifier les descripteurs de texture sont faites jusqu'à présent.

A notre connaissance, la première tentative date vers la fin des années 1970, elle a été proposée par *Haralick* (1979) [230]. Ce dernier a divisé les descripteurs de texture en *statistiques* et *structurels*, mais il a rapidement reconnu qu'il est très difficile de tracer une frontière nette entre les deux classes [231,232]. Cette division a été inspirée par le travail

pionnier de *Julesz* (1975) [233], qui a conjecturé que la discrimination de texture dans le système visuel humain existe sous deux formes: *perspective* et *cognitive*. La première fournit une caractérisation immédiate de la texture qui est essentiellement statistique, tandis que la dernière nécessite un examen qui est généralement structurel.

Wu et Chen (1992) [234] ont affiné cette taxonomie de deux classes en divisant la classe des méthodes statistiques en cinq sous-classes: méthodes spatiales dépendante au niveau de gris, caractéristiques spatiales basées sur les fréquences, caractéristiques basées sur des modèles stochastiques, méthodes de filtrage, et approches heuristiques. Puis, vers la fin des années 1990, *Tuceryan et Jain* (1998) [226] ont proposé une classification à quatre catégories (méthodes : statistiques, géométriques, basées sur le modèle, et de traitement du signal) qui a recueilli un bon nombre de palpeurs et a fortement influencé la littérature. Pourtant, la classification a été exempte de critiques, certains descripteurs de texture possèdent des traits très distinctifs qui appartiennent à plus d'une classe, et donc en général une séparation complètement croquante ne se tient pas. Récemment, *Xie et Mirmehdi* (2008) [235] ont suggéré que les quatre classes proposées par *Tuceryan et Jain* doivent être plutôt considérées comme des attributs qu'une méthode spécifique peut les posséder ou non. Une telle catégorisation représente, à notre avis, la meilleure tentative pour classifier les descripteurs de texture. Pourtant, toute classification fondée sur des catégories "*sémantiques*" ne sera jamais satisfaisante, en raison de sa nature intuitive et informelle. Au contraire, l'approche correcte doit être basée sur des définitions "*mathématiques formelles*".

Les difficultés mentionnées ci-dessus sont soulevées clairement quand il s'agit de trouver le bon placement pour la méthode LBP et ses variantes relatives. Bien que le LBP a été proposé comme « une approche d'unification pour les modèles statistiques et structurels traditionnellement divergents dans l'analyse de texture » [236], il n'existe actuellement aucun consensus sur ce point, en raison de l'absence d'une taxonomie universellement acceptée. Plusieurs auteurs classent le LBP de différentes manières comme: purement statistique [237], purement structurel [238], stochastique [239], ou même à base de modèle [240].

4.3.3 Problèmes d'analyse de texture

Lors de l'application d'une l'analyse de texture dans un environnement réel, selon les mesures exigées, différents sous-problèmes d'analyse de texture sont rencontrés. *Tuceryan et Jain* (1998) [226] ont listé quatre problèmes d'analyse de texture: la classification de texture,

la segmentation de texture, la synthèse de texture, et la détermination d'une forme par texture. Par contre, *Petrou et Sevilla* (2006) [229] ont fait une liste de trois problèmes comprenant: la classification de texture, la segmentation de texture, et la détection des défauts par texture. Dans ce qui suit, ces problèmes sont discutés en plus de détails.

a. Classification de texture

Dans la classification de texture, le but consiste à assigner un échantillon de texture inconnu dans l'une des classes prédéfinies (Figure 4.2). L'attribution se fait sur la base de règles qui sont généralement dérivées automatiquement à partir d'un ensemble d'apprentissage composé par des échantillons de texture avec des classes connues.

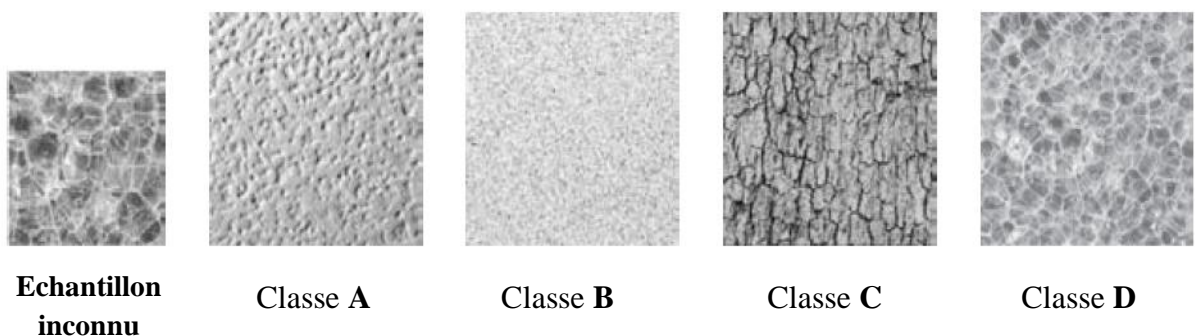


Figure 4. 2 : Exemple de classification de texture.

Étant donné une image de texture segmentée pour être classifiée, les deux composantes essentielles sont l'extracteur de caractéristiques et l'algorithme de classification. Un aperçu sur les approches d'extraction des caractéristiques de texture est présenté dans la sous-section suivante (4.3.4). Pour une introduction générale et un examen sur les classificateurs de motifs statistiques, voir les références [241,242], ou pour un regard critique sur les progrès dans la recherche sur les classificateurs, voir la référence [243].

Dans la classification de texture, le classificateur de type k-NN avec différentes mesures de distances est considéré comme un choix *commun* [244,245]. Récemment, le SVM a gagné plus d'intérêt, les résultats rapportés avec ce classificateur ont surpassé les résultats du k-NN [246,247].

Le problème de la récupération de texture dans une certaine mesure est lié à la classification de texture. Essentiellement, la récupération de texture est une opération basée sur le contenu de l'image appliquée aux images de texture. Ainsi, l'objectif est de récupérer à partir d'une base de données autant d'échantillons de textures demandées que possible.

Cependant, la texture est plus souvent utilisée comme une caractéristique additionnelle pour une récupération générale de l'image.

b. Segmentation de texture

Dans la segmentation de texture, le but consiste à diviser une image en régions cohérentes par l'utilisation de l'information de texture (Figure 4.3). Dans la segmentation supervisée de texture, le système dispose des modèles de textures pour être rencontrés dans les images à segmenter. D'autre part, la segmentation non-supervisée de texture a pour objectif de diviser une image en régions de textures similaires sans aucune information a priori sur les différentes textures. La segmentation de l'image, quoiqu'un problème mal posé, elle a plusieurs applications pratiques, ainsi que la texture s'est avérée comme un signal utile dans la segmentation.

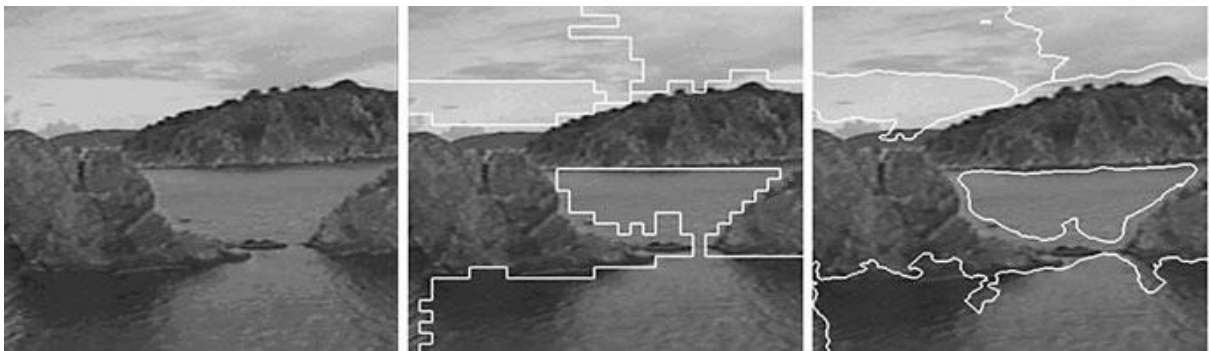


Figure 4. 3 : Exemple de segmentation de texture.

La détection des défauts par texture est un sous-problème de segmentation de texture, elle est généralement rencontrée dans l'inspection visuelle. Dans ce problème, nous avons un modèle de texture "*acceptable*", et l'objectif consiste à analyser l'image de texture pour trouver des défauts, qui sont des événements locaux habituellement dérivés par un modèle.

c. Détermination d'une forme par texture

Le problème de la détermination d'une forme par texture consiste à déduire la forme de trois dimensions (3D) d'un objet à partir de son image. Il est prouvé que la texture est un repère important dans la perception de la forme 3D par les humains [248]. Parmi les recherches dans le domaine de la vision par ordinateurs, l'une des stratégies récentes proposée pour la détermination d'une forme par texture qui consiste à modéliser la déformation des éléments de textures individuelles, comme proposée par *Lobay et Forsyth* (2006) [249].

Cependant, ils précisent dans leur article que « les applications pour la détermination d'une forme par texture sont largement absentes ».

d. Synthèse de texture

L'objectif de la synthèse de texture est de synthétiser plusieurs échantillons de textures similaires de manière perceptuelle. Après plusieurs années de recherches dans le domaine de synthèse de textures, par application de contraintes statistiques sur les images de sortie [250,251], les meilleurs résultats semblent actuellement être produits par les approches basées sur les patches (pièces) de l'image, suggérées pour la première fois dans la synthèse de texture par *Efros et Freeman* (2001) [252].

4.3.4 Description de texture

Différentes manières pour regrouper les modèles de texture sont discutées dans la littérature de l'analyse de texture. Ces taxonomies sont rarement approfondies, car certains modèles peuvent avoir des propriétés de plusieurs groupes et quelques modèles semblent n'appartenir à aucun groupe. Ces regroupements sont toujours utiles pour aider à comprendre la variété de différents modèles.

Comme nous l'avons discuté dans la section 4.3.2, la classification proposée par *Tuceryan et Jain* (1998) [226] divise les modèles de texture en quatre groupes: méthodes *statistiques* telles que les matrices de cooccurrences et les caractéristiques d'auto-corrélation, méthodes *géométriques* basées sur l'analyse des propriétés géométriques des primitives de la texture, méthodes *basées sur le modèle* visant à fournir un prototype pouvant être utilisé à la fois à la description et à la synthèse de textures et méthodes de *traitement du signal* utilisant typiquement quelques mesures d'énergie à partir des images de texture filtrées.

En imagerie, les descripteurs d'apparence de l'image peuvent être considérés comme des descripteurs généraux de l'image, mais la plupart des descripteurs globaux proviennent effectivement du champ d'analyse de texture. Autrement dit, ils sont initialement proposés pour la classification ou la segmentation de texture.

L'utilisation des filtres de *Gabor* dans l'analyse de texture remonte aux années 1980. *Turner* (1986) [253] a étudié la discrimination de différentes textures par les réponses du filtre de *Gabor*. Plus tard, *Bovic et al.* (1990) [254] ainsi que *Jain et Farrokhnia* (1991) [255] ont appliqué les filtres de *Gabor* pour la segmentation non-supervisée de texture. Jusqu'ici, une

grande variété de recherches à l'aide des filtres de *Gabor* dans différents problèmes d'analyse de texture a été stimulée. L'un des meilleurs descripteurs basé sur les filtres de *Gabor* est celui développé par *Manjunath et Ma* (1996) [256], où les caractéristiques de texture sont obtenues par le calcul des *moyens* et des *écarts-types* des réponses du filtre de *Gabor* sur une image de texture. Pour un bref aperçu et comparaison sur les différentes approches de description de texture basées sur les filtres de *Gabor*, voir la référence [257].

L'une des issues les plus récemment découvertes dans la description de texture, est le problème posé par les surfaces de trois dimensions. La texture perçue par ces surfaces n'est pas causée par des changements dans l'albédo de la surface seulement, mais aussi par l'auto-ombrage et l'auto-occlusion possibles provoqués par des petites variations d'échelle dans la forme de la surface. Cela cause des variations, dues à l'illumination ou aux changements de l'angle de visionnement, qui font que la texture devienne complexe et donc difficile à modéliser ou manipuler par un descripteur de texture.

Afin de restaurer le regroupement pour les sorties des filtres, *Leung et Malik* (2001) [258] ont proposé un descripteur de texture 3D qui enregistre précisément un ensemble d'images d'une texture par la convolution de chaque image enregistrée avec une banque de filtre. Ensuite, les prétendus *textons 3D* peuvent être formés par le regroupement des vecteurs comprenant des réponses de chaque filtre à la même position de chaque image enregistrée. L'histogramme des *textons 3D* est ensuite utilisé comme un descripteur.

Une stratégie récemment proposée pour la description de texture, consiste à prendre une représentation basée sur le patch (la pièce) local(e) et de l'appliquer à la description de texture. Cette approche a été proposée pour la première fois par *Lazebnik et al.* (2005) [259], qui ont présenté une méthode de représentation de texture en utilisant des détecteurs de la région d'intérêt et en regroupant les descripteurs locaux calculés au sein de ces régions.

4.3.5 Descripteurs de texture locaux

L'une des approches de description de texture qui a récemment gagné beaucoup d'intérêt est celle basée sur les descripteurs locaux de l'image. Contrairement aux descripteurs globaux où les caractéristiques codifient l'apparence de l'image entière, les descripteurs locaux considèrent les petites sous-régions de l'image.

Selon le contexte de l'application, les régions locales à décrire peuvent être obtenues autour de points d'intérêts automatiquement détectés [260] ou de régions covariantes affinées [261]. L'échantillonnage aléatoire a été proposé par *Nowak et al.* (2006) [262] avant l'échantillonnage dense, typique dans la description de texture, qui a été appliqué par *Tuytelaars et Schmid* (2007) [263]. Une grille fixe utilisée pour déterminer les régions locales a été proposée par *Ahonen et al.* (2006) [264] et *Vogel et Schiele* (2007) [265].

Comme le patch local à décrire, probablement transformé en une orientation canonique, est déterminé, il existe une variété d'options pour construire les caractéristiques photométriques. Nous pouvons appliquer certains descripteurs globaux sur chaque patch local de l'image ou employer des descripteurs spécifiquement conçus pour les régions locales. Une comparaison entre les descripteurs locaux, faite par *Mikolajczyk et Schmid* (2006) [266], a évalué plusieurs descripteurs locaux; ils ont conclu que l'histogramme de l'orientation locale du gradient (GLOH pour Gradient Location-Orientation Histogram) [267] fournit les meilleurs résultats, suivi du descripteur de la transformée de caractéristiques visuelles invariantes à l'échelle (SIFT) [196]. Ces deux descripteurs sont basés sur le calcul des histogrammes de l'orientation du gradient dans les sous-régions du patch à décrire. Les caractéristiques robustes et accélérées (SURF) sont un descripteur local introduit récemment par *Bay et al.* (2006) [202]. Le descripteur SURF est basé sur les *ondelettes de Haar*; il est rapide à calculer. Il a été également déterminé que SURF présente des résultats meilleurs que SIFT, en dépit de sa plus petite dimension de description [202].

Le descripteur de l'histogramme des gradients orientés (HOG pour Histogram of Oriented Gradients) [267] basé sur un histogramme pondéré des directions de gradient a montré de bonnes performances dans la détection humaine (ex., visage, action ou mouvement). En outre, l'opérateur de motif binaire local (LBP) a été également étendu pour satisfaire aux besoins de la région d'intérêt par *Heikkilä et al.* (2009) [268], grâce à sa puissance discriminative et à la simplicité de calcul.

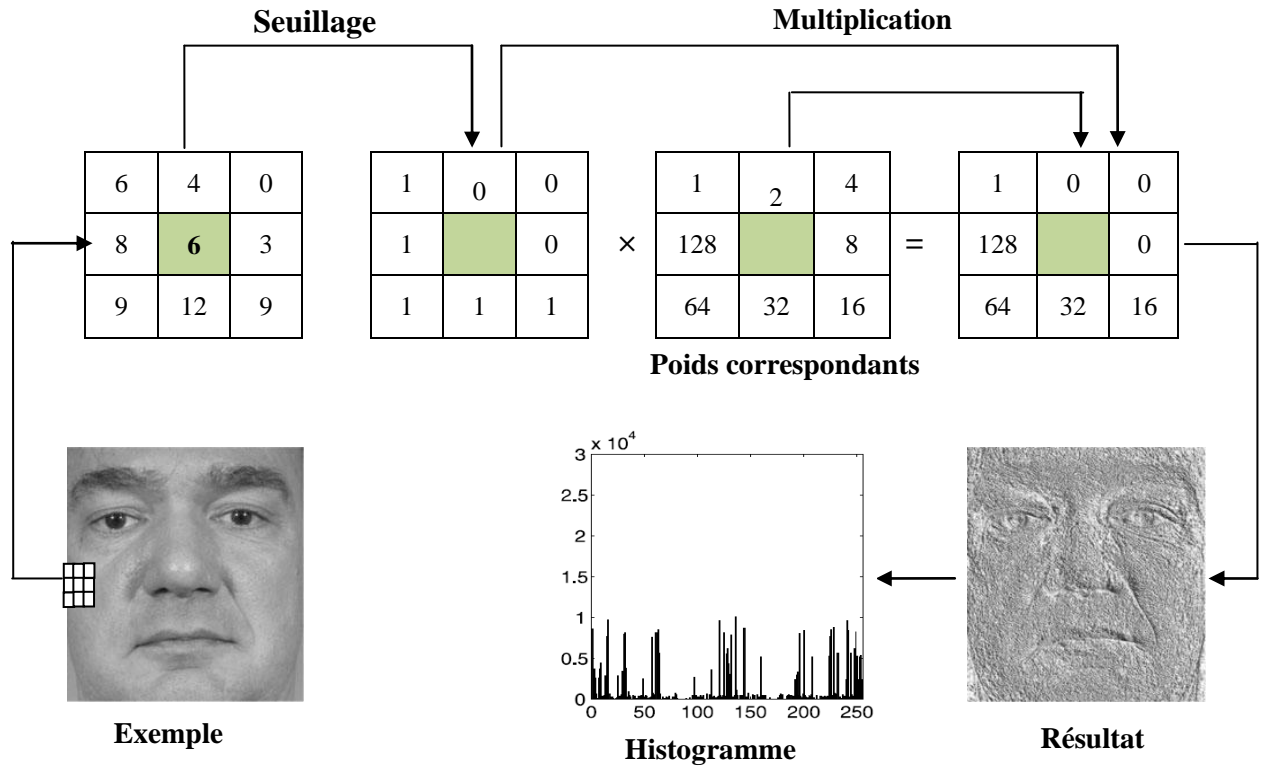
4.4 Motif binaire local (LBP: Local Binary Pattern)

Le motif binaire local est un opérateur de texture simple mais très efficace, il marque les pixels d'une image par le seuillage du voisinage de chaque pixel et considère le résultat comme un nombre binaire. La méthode LBP peut être considérée comme une approche d'unification pour les modèles statistiques et structurels traditionnellement divergents dans

l'analyse de texture. La propriété la plus importante de l'opérateur LBP dans les applications du monde réel réside dans son invariance contre les changements monotones du niveau de gris causés, par exemple, par des variations d'éclairage. Une autre propriété aussi importante réside dans sa simplicité de calcul, qui permet d'analyser des images compliquées en temps réel.

L'opérateur LBP original, introduit par *Ojala et al.* (1996) [244], est fondé sur l'hypothèse qui suggère que la texture possède localement deux aspects complémentaires: un motif et sa force. L'opérateur fonctionne dans un voisinage de 3×3 , en utilisant la valeur centrale comme un seuil. Un code LBP décrivant la texture locale d'une région est généré par: tous les voisins prendront la valeur "1" si leur valeur est supérieure ou égale au pixel courant et "0" autrement (Figure 4.4). Les pixels de ce motif binaire sont alors multipliés par des poids et sommés afin d'obtenir un code LBP du pixel courant.

Comme le voisinage est composé de 08 pixels, un total de $2^8 = 256$ labels différents peut être obtenu en fonction des valeurs grises relatives au centre et aux pixels dans le voisinage. La mesure du contraste C est obtenue par la soustraction de la moyenne des niveaux de gris au-dessous du pixel central de celle des niveaux de gris au-dessus (ou égale) du pixel central. Si tous les huit voisins seuillés du pixel central ont la même valeur (0 ou 1), la valeur du contraste est mise à 0. Les distributions des codes LBP, ou les distributions bidimensionnelles du LBP et le contraste local (LBP/C), sont utilisées comme des caractéristiques dans la classification ou la segmentation. La figure 4.4 représente le fonctionnement d'un opérateur LBP basique. Quelques images de visages et d'oreilles ainsi que leurs codes LBP correspondants sont illustrées dans la figure 4.5.



Le motif = 11111000 **LBP code** = 1 + 16 + 32 + 64 + 128 = 241 $C = \frac{6+7+9+8+7}{5} - \frac{5+2+1}{3} = 4.7$

Figure 4.4 : Exemple d'extraction des caractéristiques en utilisant l'histogramme de l'opérateur LBP basique: Application sur une image entière de visage.



Figure 4.5 : Quelques images de visages et d'oreilles ainsi que leurs codes LBP correspondants.

Nous obtenons donc pour toute l'image, des pixels dont l'intensité se situe entre 0 et 255 comme dans une image ordinaire à 8 bits. Plutôt que de décrire l'image par la séquence des motifs LBP, on peut choisir comme descripteur de texture un histogramme de dimension 255.

Dans sa forme actuelle (décrite dans la sous-section suivante), le LBP est tout à fait différent à la version basique: la version originale a été étendue par des voisinages circulaires arbitraires, ainsi qu'un certain nombre d'extensions ont été développées. L'idée fondamentale est cependant identique: le voisinage de chaque pixel est binarisé par l'utilisation d'un seuil.

Grâce à sa puissance discriminative et sa simplicité de calcul, l'opérateur de texture LBP est devenu une approche très populaire dans diverses applications. Le grand succès du LBP dans plusieurs problèmes d'analyse de texture a montré que les banques de filtres avec grandes zones de couverture ne sont pas nécessaires pour un rendement élevé dans la classification de texture, par contre les opérateurs définis pour des petits voisinages comme LBP sont généralement adéquats. Les résultats récents montrent que les approches basées sur le LBP ont un potentiel important pour de nombreuses tâches importantes en vision par ordinateurs et qui ne sont même pas considérées comme des problèmes de texture. Une bonne exploitation de l'information de texture peut augmenter d'une manière très significative la performance et la fiabilité de nombreuses tâches et systèmes de vision artificielle, aidant à rendre la technologie intrinsèquement robuste et simple à utiliser dans le monde réel.

L'opérateur LBP transforme une image dans un tableau ou une image de labels (étiquettes) entiers décrivant l'apparence à petite échelle de l'image. Ces labels ou leurs statistiques, le plus souvent l'histogramme, sont ensuite utilisés pour une analyse approfondie de l'image. Les versions les plus répandues sont conçues pour des images monochromes fixes, aujourd'hui elles sont prolongées pour traiter des images en couleur (multicanaux) ainsi que des vidéos et des données volumétriques.

4.4.1 Dérivation de l'opérateur LBP générique

Quoique l'aspect essentiel de la texture réside probablement dans les relations inter-pixels relatives aux pixels proches entre eux, les informations sur les relations entre les pixels seulement voisins ne capturent pas toujours l'essentiel de la texture. Après plusieurs années de sa publication originale, l'opérateur LBP a été présenté sous une forme révisée et plus générique par *Ojala et al.* (2002) [269]. Contrairement à la version basique du LBP qui utilise

une banque de 08 pixels dans un bloc de 3×3 , cette formulation générique de l'opérateur ne met aucune limitation sur la taille du voisinage ou le nombre de points d'échantillonnage [236,269,270].

L'opérateur LBP générique utilise des voisinages circulaires et une interpolation bilinéaire des valeurs de pixels, englobant n'importe quel rayon et nombre de pixels dans le voisinage. Dans ce qui suit, la notation (P, R) est utilisée pour les voisinages du pixel, ce qui signifie: les points d'échantillonnage P sur un cercle de rayon R . La figure 4.6 présente des exemples de voisinages circulaires.

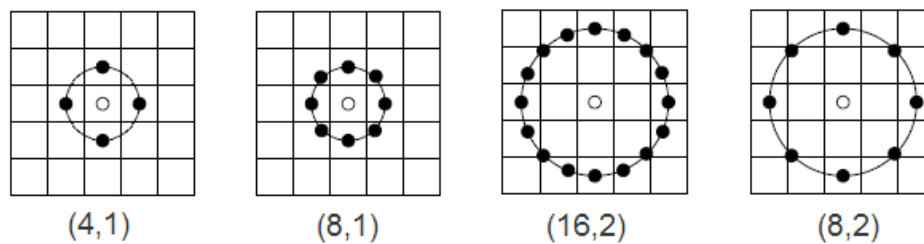


Figure 4. 6 : Exemple de voisinages (4,1), (8,1), (16,2), et (8,2) circulaires.

Étant dénoté la valeur du pixel central (x, y) par $t_c = I(x, y)$ et les valeurs grise P des points d'échantillonnage par: $t_1 = I(x_1, y_1), t_2 = I(x_2, y_2), \dots, t_p$, où les points d'échantillonnage se situent aux coordonnées $(x_p, y_p) = (x + R \cos(2\pi p/P), y - R \sin(2\pi p/P))$. L'opérateur $LBP_{P,R}$ générique est défini comme:

$$LBP_{P,R} = \sum_{p=1}^P s(t_p - t_c) 2^{p-1} \quad (4.1)$$

Où

$$s(z) = \begin{cases} 1, & \text{si } z \geq 0 \\ 0, & \text{sinon} \end{cases} \quad (4.2)$$

Une autre extension de l'opérateur original utilise ce que nous appelons: *des modèles uniformes* [269]. Pour cela, une mesure d'uniformité d'un motif est utilisée: $U(\text{"motif"})$ correspond au nombre de transitions de bits du 0 à 1, ou vice versa, quand le motif binaire est considéré comme circulaire. Un LBP est appelé *uniforme* si sa mesure d'uniformité est au maximum égale à 2. Par exemple, les motifs 00000000 ($U = 0$), 01110000 ($U = 2$), et 11001111 ($U = 2$) sont uniformes, tandis que les motifs 11001001 ($U = 4$) et 01010010 ($U = 6$) ne le sont pas. Dans le mappage du LBP uniforme, il y a un label de sortie séparé pour chaque motif uniforme, et tous les motifs non-uniformes sont assignés à un seul label.

Comme les motifs du $LBP_{P,R}$ sont obtenus par échantillonnage circulaire autour du pixel central, la rotation de l'image en entrée a deux effets: chaque voisinage local est tourné dans un autre emplacement de pixels, et dans chaque voisinage, les points d'échantillonnage sur le cercle entourant le point central sont tournés dans une orientation différente (voir Figure 4.7). Par conséquent, les labels LBP invariables à la rotation peuvent être obtenus en appliquant un mappage qui fait tourner circulairement chaque motif binaire à la valeur minimale. Par exemple, les séquences binaires 1000011, 1110000, et 0011100 proviennent de différentes rotations du même motif local, et elles correspondent toutes à la séquence normalisée 0000111.

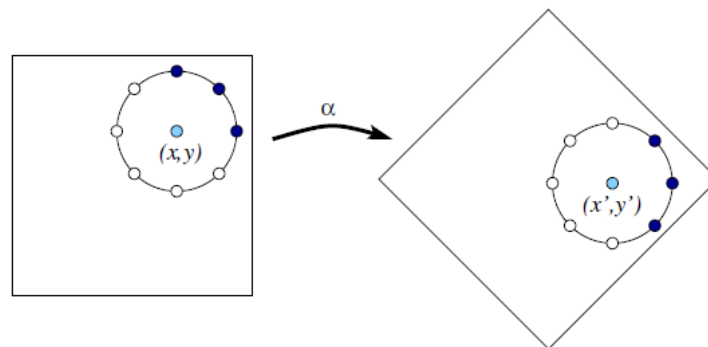


Figure 4. 7 : Effet de rotation de l'image sur les points d'échantillonnage dans un voisinage circulaire.

4.4.2 Utilisation du LBP en caractérisation biométrique

Dans l'utilisation du LBP pour la classification de texture, les cooccurrences des codes LBP dans une image sont collectées dans un histogramme. La classification est ensuite effectuée par le calcul des similitudes simples de l'histogramme. Cependant, étant considéré une approche similaire pour la représentation, par exemple, d'une image faciale, il résulte malheureusement une perte dans l'information spatiale, donc il faut codifier les informations de texture tout en préservant également leurs emplacements. Une technique simple pour atteindre cet objectif consiste à utiliser les descripteurs de texture LBP pour construire plusieurs descriptions locales du visage et les combiner en une description globale. Ces descriptions locales ont récemment gagné plus d'intérêts, elles sont compréhensibles étant donné les limitations des représentations holistiques. Ces méthodes basées sur les caractéristiques locales semblent être plus robustes contre les variations de pose ou d'illumination par rapport aux méthodes holistiques.

Une autre raison pour choisir une approche basée sur les caractéristiques locales est que l'essai de construire une description globale d'un visage à l'aide des méthodes de texture n'est pas raisonnable, puisque les descripteurs de texture tendent à faire la moyenne sur la région de l'image. C'est une propriété souhaitable pour les textures, parce que la description de texture doit être habituellement invariante à la traduction ou même la rotation, surtout pour les petites textures répétitives; les relations à petite échelle déterminent l'apparence de la texture et donc les relations à grande échelle ne contiennent pas des informations utiles. Pour les visages, cependant, la situation est différente: la conservation des informations sur les relations spatiales est importante.

La méthodologie de base pour la description d'une modalité biométrique (ex., visage ou oreille) à l'aide du LBP se présente comme suit: l'image de la modalité biométrique est divisée en régions locales et les descripteurs de texture LBP sont extraits de chaque région indépendamment. Les descripteurs sont ensuite concaténés pour former une description globale de la modalité, comme montré dans la figure 4.8.

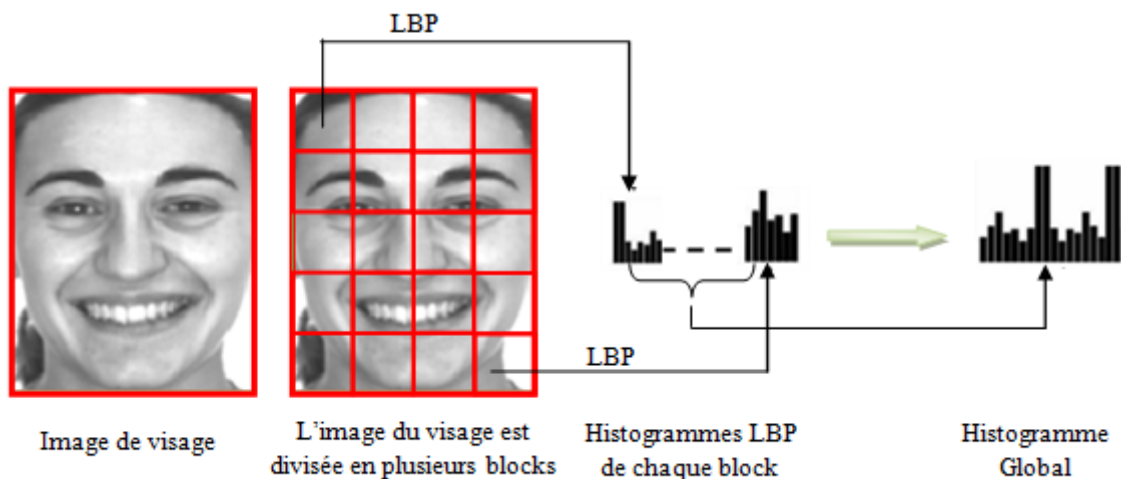


Figure 4.8 : Exemple de la représentation d'une modalité biométrique (visage) à base de LBP.

L'histogramme de base utilisé pour recueillir des informations sur les codes LBP dans une image peut être étendu dans un histogramme spatialement amélioré qui codifie à la fois l'apparence et les relations spatiales sur les régions de la modalité biométrique. Comme les régions de la modalité R_0, R_1, \dots, R_{m-1} sont déterminées, l'histogramme spatialement amélioré est défini comme:

$$H_{i,j} = \sum_{x,y} I\{f_l(x,y) = i\} I\{(x,y) \in R_j\}, i = 0, \dots, n-1, j = 0, \dots, m-1 \quad (4.3)$$

Cet histogramme présente effectivement une description de la modalité à trois niveaux différents de localité: les labels du LBP pour l’histogramme contiennent des informations sur les motifs au niveau pixel, les labels sont additionnés sur une petite région pour produire l’information au niveau régional, et les histogrammes régionaux sont concaténés pour construire une description globale de la modalité.

Il est convenu de noter que: lorsque nous utilisons des méthodes basées sur l’histogramme, les régions R_0, R_1, \dots, R_{m-1} ne sont pas nécessairement; elles ne doivent pas être de la même taille ou forme, et ne doivent pas nécessairement couvrir l’image entière. Il est également possible d’avoir des régions partiellement chevauchantes.

4.5 LBP unidimensionnel (1DLBP)

Une nouvelle variante de la technique LBP a été proposée dans notre laboratoire PIMIS et en collaboration avec le laboratoire PRISME de l’université d’Orléans (France) par *Houam et al.* (2013) [271]. Cette variante a été initialement utilisée pour la caractérisation des textures osseuses; contrairement aux techniques présentées précédemment, cette méthode a été appliquée d’une manière unidimensionnelle sur des projections linéaires, ainsi l’histogramme calculé a été utilisé pour discriminer les textures osseuses [272,273]. Par conséquent et en continuation à ces travaux, nous avons proposé, dans le cadre de cette thèse, d’exploiter les avantages de ce descripteur pour la caractérisation de données biométriques.

Le choix de l’approche unidimensionnelle, notée 1DLBP (1D Local Binary Pattern) est motivée par le fait que les projections linéaires sur un espace 1D (passage du 2D vers 1D) d’une image texturée [273] présentent des informations utiles pour la codification des motifs globaux et locaux de la texture [274-276]. Le concept de cette méthode consiste à décrire dans un code binaire l’agitation locale d’un signal 1D; un masque linéaire W est construit pour balayer séquentiellement l’image projetée avec un pas d’un élément. W est un nombre impair assurant le voisinage symétrique de chacun des éléments du masque, $m = W - 1$ représente le nombre de voisins concernés, t_c est l’élément central du masque 1DLBP, tandis que t_0, t_1, \dots, t_{m-1} sont les voisins de l’élément central t_c . La figure 4.9 présente les quatre ensembles linéairement symétrique de voisins possibles pour différentes valeurs de m et W .

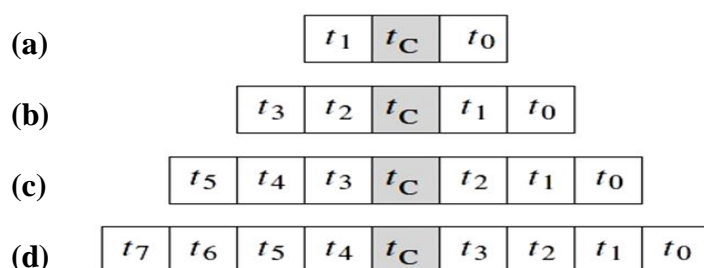


Figure 4. 9 : Ensemble de voisins linéairement symétrique pour différentes valeur de m et W .

$(m, W) =$ (a) (3,2), (b) (5,4), (c) (7,6), ou (d) (9,8).

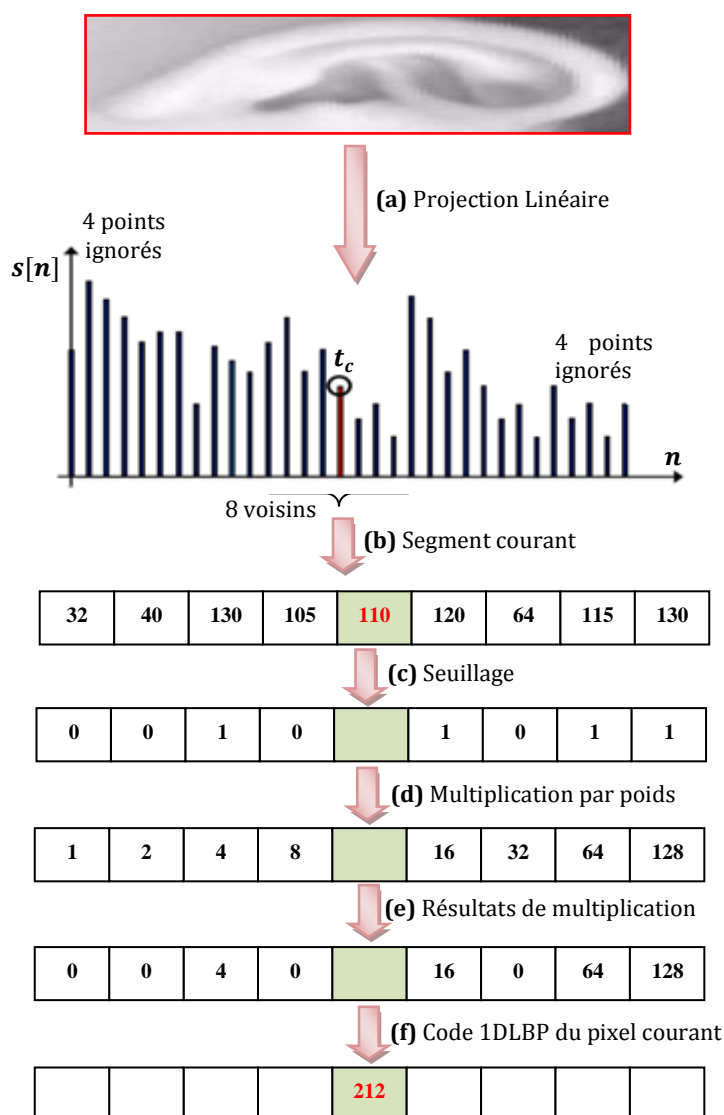


Figure 4. 10 : Application du 1DLBP sur une image d'oreille normalisée (masque de 08 voisins).

Le descripteur 1DLBP peut être utilisé avec 2, 4, 6, ou 8 voisins. Le codage est calculé par le seuillage des valeurs du voisinage avec l'élément central t_c . Tous les voisins prennent la valeur "1" s'ils sont supérieurs ou égaux à l'élément central et la valeur "0" dans le cas contraire. Ensuite, chaque élément binaire du vecteur résultant est multiplié par un poids lié à sa position dans le masque. Le nombre binaire est ordonné conventionnellement dans le vecteur de la gauche vers la droite. En conséquence, un nombre naturel appartenant à l'intervalle $[0, 2^m - 1]$ est généré. Enfin, la valeur de l'élément courant est remplacée par la somme du vecteur résultant. Ceci peut être récapitulé par:

$$1DLBP_{m,W} = \sum_{p=1}^{m-1} s(t_p - t_c) \times 2^p \quad (4.4)$$

Où

$$s(z) = \begin{cases} 1, & \text{si } z \geq 0 \\ 0, & \text{sinon} \end{cases} \quad (4.5)$$

Où t_c et t_p sont respectivement les valeurs de l'élément central et ses voisins dans le masque 1D. L'indice p augmente de la gauche vers la droite (Figure.4.9.d). Un exemple d'un masque de motifs 1×8 et ses codes binaires appliqués sur une image d'oreille normalisée est présenté dans la figure 4.9.

Le descripteur 1DLBP est défini par l'histogramme des motifs 1D qui est calculé par:

$$H_k = \sum_{\frac{W}{2} \leq n \leq N - \frac{W}{2}} \delta(1DLBP_{m,W}(s[n], k)) \quad (4.6)$$

Où H_k est le descripteur 1DLBP du signal 1D projeté, k est le nombre de barres ($k = 2^{W-1}$), δ est l'opérateur de *Kroniker*, W est la taille du masque 1DLBP, m est le nombre de voisins, et N représente la taille du signal projeté.

4.6 Quantification par phase locale (LPQ: Local Phase Quantization)

Pour aborder la sensibilité relative du LBP au flou, un descripteur de type *Quantification par Phase Locale* (LPQ pour Local Phase Quantization) a été proposé par *Ojansivu et Heikkil* (2008) [277], ce descripteur est basé sur la quantification de la phase de la transformée de *Fourier* dans les voisinages locaux. La phase peut être vue comme un flou invariant sous certaines conditions généralement remplies. Dans l'analyse de texture, les histogrammes des labels LPQ calculés dans des régions locales sont utilisés en tant que descripteur de texture

d'une manière similaire à la méthode LBP. Le descripteur LPQ a récemment reçu un large intérêt pour la reconnaissance de visages en flou-invariant. Le LPQ est insensible aux images floues, et il s'est avéré être un descripteur très efficace dans la reconnaissance de visages avec des images floues ou nettes.

Dans l'approche LPQ, le flou spatial est représenté par une convolution entre l'intensité de l'image et la fonction de diffusion du point (PSF pour Point Spread Function). Dans le domaine fréquentiel, ceci se traduit par une multiplication de type: $G = F \cdot H$ où G , F , et H sont la transformée de *Fourier* de l'image floue, l'image nette, et le PSF respectivement. Etant considéré la phase du spectre seulement, la relation se transforme en une somme: $\angle G = \angle F + \angle H$. Si le PSF est centralement symétrique, la transformée H devient une valeur réelle et l'angle de phase $\angle H$ doit être égal à 0 ou π . En outre, la forme de H pour un PSF régulier est proche à une *Gaussienne* ou à une fonction *sigmoïde*, ce qui fait qu'au moins les valeurs de basses fréquences de H sont positives. A ces fréquences, $\angle H = 0$ ce qui fait que $\angle G = \angle F$ soit une être une propriété invariante au flou; ce phénomène est la base de la méthode LPQ.

Pour une image donnée, chaque pixel est marqué par un code LPQ invariant au flou. Similaire à la méthodologie du LBP, les cooccurrences de ces codes LPQ sont collectées dans un histogramme pour la classification. Une implémentation basique du LPQ est illustrée dans la figure 4.11. Quelques images de visages et d'oreilles ainsi que leurs codes LPQ basiques sont présentées dans la figure 4.12.

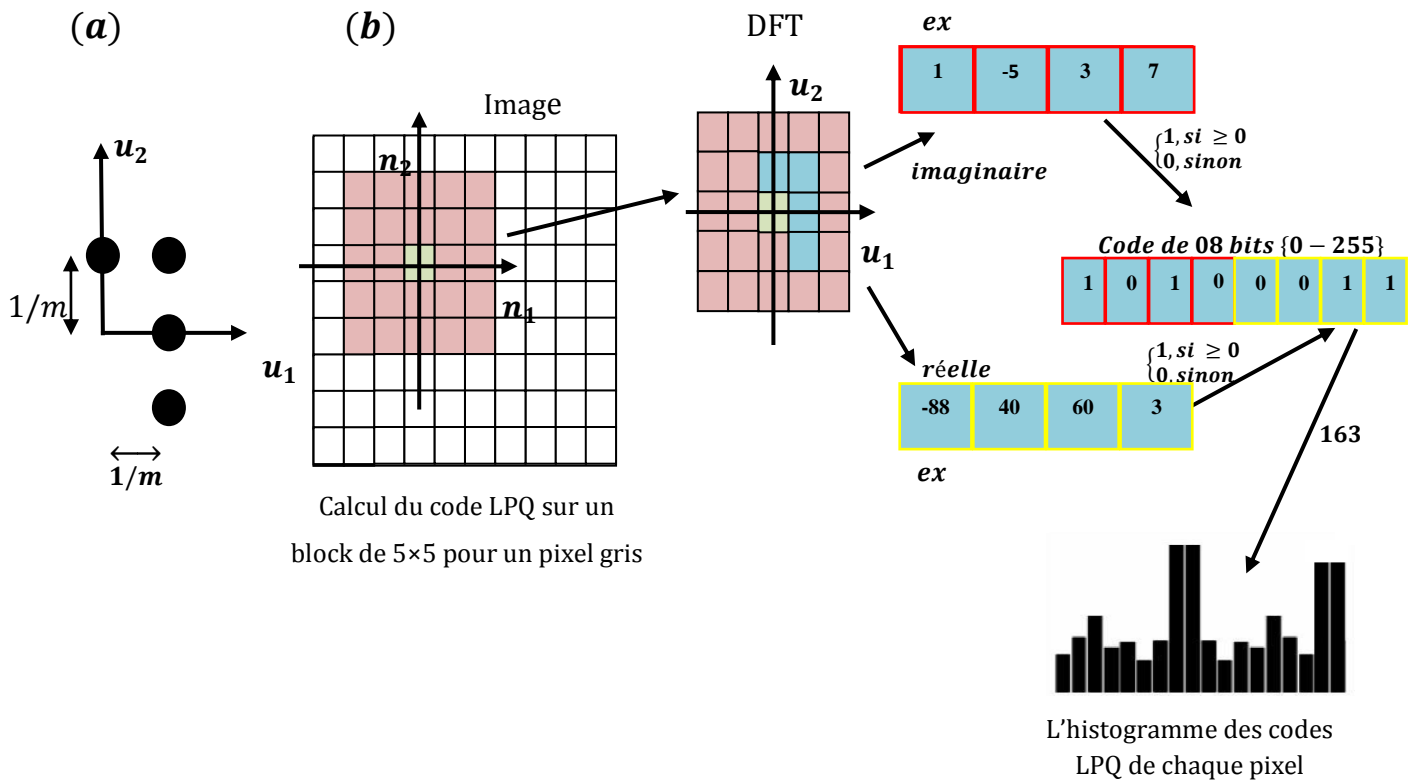


Figure 4.11 : Implémentation basique du LPQ: (a) Exemple de fréquence utilisée (b) Exemple de calcul par le descripteur.

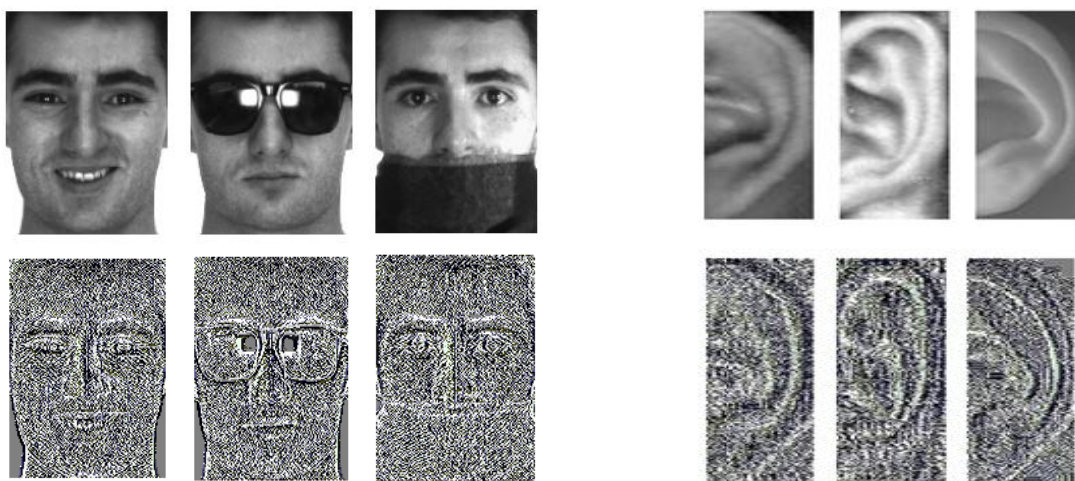


Figure 4.12 : Caractéristiques statistiques et binarisées de l'image (BSIF: Binarized Statistical Image Features)

4.7 Caractéristiques statistiques et binarisées de l'image (BSIF : Binarized Statistical Image Features)

Le descripteur BSIF a été récemment proposé par *Kannala et Rahtu* (2012) [278], il a été utilisé pour la reconnaissance de visage et la classification de texture. Basé sur LBP et LPQ, l'idée derrière le BSIF consiste à apprendre automatiquement un ensemble fixe de filtres à partir d'un petit ensemble d'images naturelles, au lieu d'utiliser des filtres fabriqués-à-la-main comme LBP ou LPQ. BSIF implique un apprentissage, au lieu d'un réglage manuel, pour obtenir une représentation statistiquement significative de l'image, qui permet d'encoder l'information efficace en utilisant la quantification par élément simple. L'apprentissage fournit également une manière facile et flexible pour ajuster la longueur du descripteur et de l'adapter aux applications présentant des caractéristiques d'images inhabituelles.

Pour caractériser les propriétés de la texture dans chaque sous-région de l'image, les histogrammes des labels BSIF sont alors utilisés. La valeur de chaque élément (bit) dans la chaîne du code binaire BSIF est calculée par la binarisation de la réponse d'un filtre linéaire, avec un seuil à zéro. Chaque bit est associé à un filtre différent et la chaîne de bits détermine le nombre de filtres utilisés. L'ensemble de filtres est appris (formé) à partir d'un ensemble de patches d'images naturelles en maximisant l'indépendance statistique des réponses du filtre.

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

$$s_i = \sum_{u,v} W_i(u, v) X(u, v) = w_i^T x \quad (4.7)$$

Où les vecteurs w et x contiennent les pixels de W_i et X .

La caractéristique binarisée b_i est obtenue par la mise de $b_i = 1$ si $s_i > 0$ et $b_i = 0$ sinon. Les filtres W_i sont appris en utilisant l'analyse en composantes indépendantes (ICA) en maximisant l'indépendance statistique de s_i .

Le descripteur BSIF possède deux paramètres qui sont: la taille du filtre l et la longueur n de la chaîne binaire. Les filtres originaux proposés par *Kannala et Rahtu* (2012) [278] ont été appris avec 50 000 patches d'images. Les filtres obtenus avec $l = 7$ et $n = 8$ sont illustrés dans la figure 4.13. Quelques images de visages et d'oreilles ainsi que leurs codes BSIF (avec $l = 7$ et $n = 8$) sont présentées dans la figure 4.14.

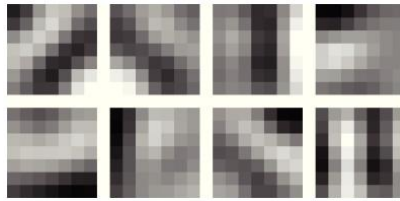


Figure 4.13 : Banque de filtres BSIF de taille 7×7 et quantification de 8bits appris par des images naturelles.

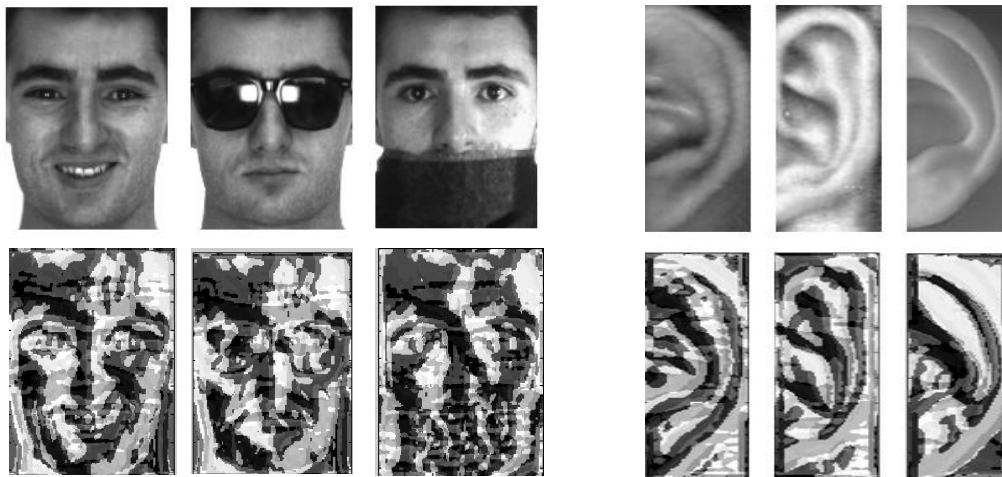


Figure 4.14 : Quelques images de visages et d'oreilles ainsi que leurs codes BSIF (avec $l = 9$ et $n = 8$) correspondants.

Contribution (idée proposée): Parmi les limitations de l'opérateur BSIF, nous avons trouvé que les filtres originaux de ce descripteur ont été entraînés à partir d'un ensemble naturel et aléatoire composé de 50 000 patches d'images texturées. Pour être plus précis, nous avons proposé, dans le cadre de cette thèse, de refaire l'apprentissage des filtres de ce descripteur, mais, uniquement par des images de visages et oreilles respectivement, et non pas avec un ensemble d'images naturelles, aléatoires, et texturées. Par conséquent, les nouveaux pourcentages de la classification obtenus, présentés dans le chapitre suivant, se sont améliorés d'une manière très significative.

L'apprentissage des filtres W est effectué d'une manière non-supervisée en utilisant l'analyse en composantes indépendantes (ICA) [279]. La version classique d'un modèle ICA peut être exprimée comme:

$$x = As \tag{4.8}$$

Où $x = (x_1, x_2, \dots, x_n)^T$ est le vecteur de variables aléatoires observées, le vecteur de variables latentes indépendantes est dénoté par $s = (s_1, s_2, \dots, s_n)^T$ (les composantes indépendantes), tandis que A est une matrice constante inconnue, dénotée par *la matrice de mélange*. Le nombre de composantes indépendantes est arbitraire et dans la plupart du temps il est affecté par le nombre d'échantillons. Si nous dénotons les colonnes de la matrice A par a_{ij} , le modèle peut être décrit par:

$$x = \sum_{i=1}^n a_i s_i \quad (4.9)$$

L'objectif de l'ICA est d'apprendre la décomposition dans l'équation 4.8 d'une manière non-supervisée, c'est-à-dire nous observons seulement x et nous voulons estimer A et s . Après la détermination de A et s , nous pouvons déduire les filtres appris W en utilisant la propriété:

$$W = A^{-1} \quad (4.10)$$

Pour notre analyse ICA, nous avons utilisé le toolbox Matlab: "*FastICA*". Les données que nous avons utilisées pour entraîner les filtres W se composent d'une collection de patches d'images de visages ou d'oreilles (selon le cas) prises de différentes bases d'images ou d'images collectées sur Internet. Nous avons alimenté une matrice x comme une entrée pour l'algorithme *FastICA*, où chaque colonne représente un patch (partie d'une image) et chaque ligne représente une variable aléatoire.

Contribution (analyse proposée): Les descripteurs LBP, 1DLBP, LPQ, et BSIF présentés dans ce chapitre sont intrinsèquement différents, mais, ils partagent fondamentalement une structure de calcul commune. Pour la bonne perfection, nous avons évalué la performance de ces descripteurs en termes de complexité algorithmique, le nombre d'opérations par pixel, et le temps d'exécution sur un processeur à usage général en termes de complexité algorithmique (en analysant comment le temps de calcul croît avec la taille de l'entrée). LBP, 1DLBP, LPQ, et BSIF sont essentiellement des opérateurs à base de pixel qui se mesurent par la taille de l'image. Etant donné que le temps de calcul se mesure avec la proportionnalité directe du nombre de pixels dans l'image d'entrée, nous pouvons conclure que les quatre descripteurs ont la même complexité algorithmique de $O(n)$, également appelée complexité linéaire, où n représente le nombre total de pixels à traiter.

Ceci signifie, en termes de complexité algorithmique; qu'il n'est pas important si une application exige l'utilisation de plusieurs images à faible résolution (ou patches) ou justes

quelques-unes en haute définition, dans la mesure où le nombre total de pixels reste le même. Puisque la complexité de ces descripteurs est la même, le temps de calcul est déterminé principalement par le nombre d'opérations par pixels requis pour chaque descripteur. Comme les filtres générés statistiquement du BSIF exigent un calcul de convolution pleine, les filtres du LBP, 1DLBP, et LPQ conçus manuellement ont des particularités de conception qui les rendent aptes à des calculs simplifiés.

4.8 Conclusion

Dans ce chapitre, nous avons présenté les principales notions utilisées dans le domaine de l'analyse de texture, qui est un problème très difficile. Cette difficulté est due essentiellement au fait qu'il n'existe pas de définition précise et rigoureuse de la notion de texture. Le choix des attributs de texture d'une manière générale est un point délicat, puisqu'il dépend de plusieurs facteurs. Les attributs sont donc à choisir avec précaution car non seulement ils dépendent de l'application considérée, mais ils influent également sur la performance de la discrimination.

Les descripteurs locaux de l'image font l'épine dorsale des approches actuelles pour la reconnaissance visuelle d'objets. La fonction du descripteur est de convertir les informations au niveau-pixel en une forme utile, qui capture les contenus les plus importants de l'image. Contrairement aux descripteurs globaux qui calculent les caractéristiques directement à partir de l'image entière, les descripteurs locaux, considérés comme les plus efficaces dans les conditions réelles, représentent les caractéristiques en petits patches locaux de l'image.

Dans le cadre de cette thèse, nous avons étudié l'utilisation de quatre descripteurs de texture locaux très récents, à savoir: LBP, 1DLBP, LPQ et BSIF, et nous avons fourni une vaste analyse comparative sur deux modalités biométriques différentes, qui sont: le visage et l'oreille, en utilisant des bases d'images appropriées. Les descripteurs ont été analysés en termes de précision de classification et complexité algorithmique. En outre, les performances de ces descripteurs sous différents paramètres sont bien discutées dans le chapitre suivant.

Chapitre 5

Implémentation et Etudes Expérimentales

5.1 Introduction

Dans le chapitre précédent, nous avons présenté la technique dite "*Motif Binaire Local (LBP: Local Binary Pattern)*" largement utilisée en caractérisation des images texturées, ainsi que ces extensions les plus populaires en analyse de texture. Nous avons aussi étudié des variantes très récentes et plus adaptées à l'analyse de texture. Ces variantes sont appelées: "*Motif Binaire Local Unidimensionnel (IDLBP: One Dimensional Local Binary Pattern)*", "*Quantification par Phase Locale (LPQ: Local Phase Quantization)*" et "*Caractéristiques Statistiques et Binarisées de l'Image (BSIF: Binarized Statistical Image Features)*".

Dans ce chapitre, nous allons tester et comparer ces descripteurs de texture récents sur des images de données biométriques, à savoir: le *visage* et l'*oreille*, afin de mettre en évidence leur performances et leur efficacités dans l'identification des individus; ces descripteurs sont comparés entre eux, en terme de pouvoir discriminant.

Pour tester les différentes approches, nous allons utiliser des méthodes de classification multi-classes supervisées, à savoir: les k-plus proches voisins (k-NN) et les séparateurs à vaste marge (SVM). L'évaluation et la comparaison sont effectuées en utilisant le mode d'identification avec des taux d'identification au rang-1, discutés également dans le premier chapitre.

5.2 Représentation et reconnaissance biométrique à base de descripteur de texture locaux

Le système biométrique proposé dans le cadre de ce travail nécessite deux phases de fonctionnement. La première est une phase d'*apprentissage*; elle consiste à enregistrer les traits de la modalité biométrique, soit visage ou oreille, de chaque individu afin de créer son propre modèle biométrique, ce dernier est alors stocké dans la base de données. La deuxième est une phase de *test*; elle consiste à enregistrer les mêmes caractéristiques et à les comparer avec les modèles biométriques stockés dans la base de données. Si les données enregistrées

correspondent à un modèle dans la base de données, l'individu, dans ce cas, est considéré comme identifié.

Soit une image de visage ou oreille, nous proposons la représentation suivante de l'image. Tout d'abord, nous effectuons un prétraitement sur l'image originale en appliquant le filtre *médian* et la normalisation de l'histogramme. L'objectif de l'étape de prétraitement consiste à préparer la représentation de l'image originale afin de faciliter la tâche des modules suivants et d'améliorer les performances de l'identification. Premièrement, l'image couleur est convertie en une image au niveau gris. Ensuite, l'image au niveau de gris est filtrée par le filtre *médian* pour réduire le bruit. A la fin, l'histogramme de l'image filtrée est normalisé pour améliorer le contraste. L'image normalisée est alors décomposée en plusieurs sous-images de différentes tailles en utilisant la projection horizontale-verticale (voir figure 5.1), afin de mieux caractériser les détails et la corrélation entre les patches de l'image. Le descripteur de texture, soit LBP, 1DLBP, LPQ ou BSIF, est ensuite appliqué sur chaque bloc (sous-image), un histogramme sera, par conséquent, extrait à partir de chaque bloc. Finalement, les histogrammes extraits de chaque bloc sont concaténés dans un histogramme de caractéristiques global représentant l'image du visage ou l'oreille. Ce dernier présente effectivement une description de la modalité à trois niveaux différents de localité: les labels du descripteur de texture local pour l'histogramme contenant des informations sur les motifs au niveau pixel, les labels sont additionnés sur une petite région pour produire l'information au niveau régional et les histogrammes régionaux sont concaténés pour construire une description globale de la modalité.

Une fois la représentation est établie, trois classificateurs, à savoir: les k-plus proches voisins (k-NN), les séparateurs à vaste marge (SVM) avec un noyau *linéaire* et avec un noyau *RBF* (Radius Basis Function) sont comparés et utilisés pour la classification. (Pour plus de détails sur ces classificateurs, voir Annexe A)

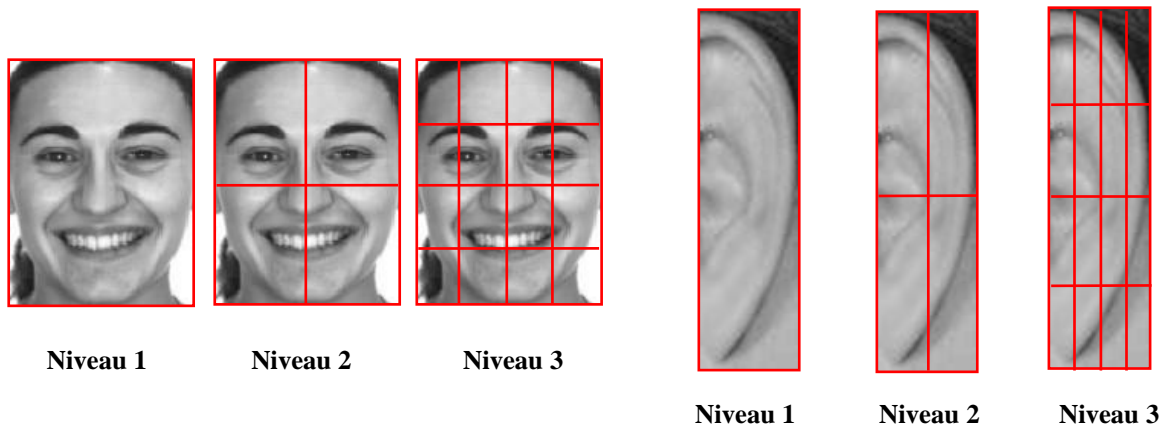


Figure 5. 1 : Stratégie de la décomposition de l'image en plusieurs sous-blocs (décomposition horizontale-verticale).

5.3 Etude expérimentale

Pour une évaluation approfondie, nous avons effectué une série d'expérimentations en utilisant les descripteurs de texture locaux, détaillés dans le chapitre précédent, à savoir: LBP, 1DLBP, LPQ et BSIF afin d'extraire des caractéristiques discriminantes à partir d'un ensemble d'images (visages / oreilles) mis à la disposition des chercheurs et fournies par les bases de données: ORL [280], AR [281], IIT Delhi [282] et USTB [283]. Nous avons donc choisi deux bases d'images pour le visage et deux autres bases d'images pour l'oreille. La description de chaque base d'images ainsi que le protocole utilisé pour évaluer les performances de chaque base de données sont discutés dans la sous-section suivante. Nous avons utilisé Matlab 2010b pour implémenter nos expérimentations sur Windows 7, HP Core 2 Duo, CPU 2.4 Ghz avec 2Gb de Ram.

5.3.1 Bases d'images

a- Le visage

Pour évaluer la robustesse des descripteurs de textures locaux (LBP, 1DLBP, LPQ et BSIF) par rapport aux légères variations de l'angle et l'alignement de la pose, nous avons testé chaque descripteur sur la base d'images de visages ORL (*Olivetti Research Laboratory Cambridge*) [280]. La base d'images contient 400 images de 40 objets distincts (personnes), 10 images pour chaque objet, prises entre Avril 1992 et Avril 1994. Des exemples d'images de la base de données ORL sont présentés dans la figure 5.2. Les images ont été prises à des moments différents pour certaines personnes, il y a des variations dans l'expression du visage

(yeux ouverts / fermés, souriant / non-souriant), les détails du visage (lunettes / sans lunettes), l'échelle (variation jusqu'à 10%). Toutes les images ont été prises sur un fond foncé et homogène avec des objets dans une position frontale et verticale, et dans une tolérance pour une inclinaison et rotation allant jusqu'à 20°. Les images sont grises avec une résolution de 92 × 112 pixels.



Figure 5. 2 : Quelques échantillons de la base d'images ORL.

Le protocole utilisé pour évaluer cette base d'images consiste à choisir aléatoirement 05 images par personne pour créer l'ensemble d'*apprentissage* et les 05 autres images restantes sont utilisées pour créer l'ensemble de *test*. Cette sélection aléatoire a été effectuée pour 100 permutations et le taux d'identification moyen a été pris en considération.

Dans le but de tester les performances des descripteurs de texture locaux, discutés ci-dessus, par rapport aux changements des expressions faciales et les circonstances de l'occlusion ainsi que la robustesse avec les bases d'images à grande échelle, nous avons testé chaque descripteur sur la base d'images de visages AR [281]. Cette base de données a été collectée au "*centre de vision par ordinateurs*" à Barcelone (Espagne) en 1998. Elle contient presque de 4000 images de visages en couleur de 134 objets (75 hommes et 59 femmes). Les images et les conditions d'enregistrement (paramètres de la caméra, configuration de l'illumination et distance de la caméra) ont été soigneusement contrôlées et constamment calibrées pour s'assurer que les configurations soient identiques pour tous les objets. Les images couleurs RGB ont la taille de 120 × 165 pixels. Elles ont été prises deux fois dans un intervalle de deux semaines. Les images de cette base de données présentent différentes expressions faciales, dans des conditions d'éclairage et d'occlusion différentes, comme montré dans la figure 5.3, ce qui augmente la difficulté en identification.

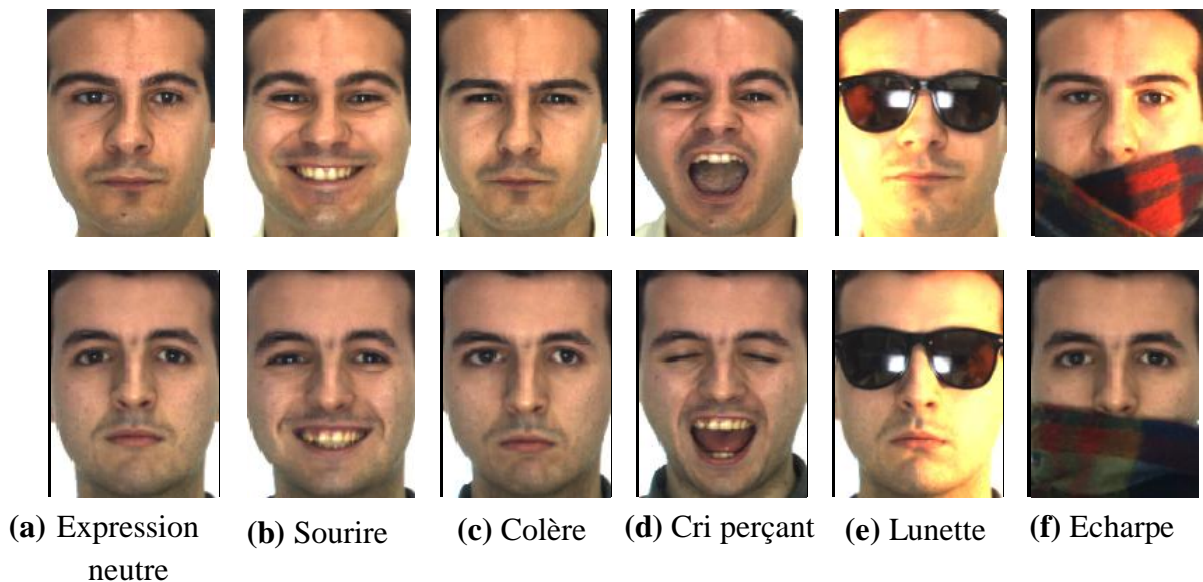


Figure 5.3 : Quelques échantillons de la base d'images AR.

Le protocole utilisé pour évaluer les performances de cette base de données (Protocole A) prend les trois premiers types d'images (c'est à dire, les images de type: expression neutre, sourire et colère) pour créer l'ensemble d'*apprentissage*, tandis que les trois types d'images restants (c'est à dire, les images de type: cri perçant, lunette et écharpe) sont utilisés pour créer l'ensemble de *test*. Un nouveau protocole B lié à cette base de données utilise les images du premier type (expression neutre) en *apprentissage* et tous les autres types (sourire, colère, cri perçant, lunette, et écharpe) sont utilisés séparément comme des ensembles de *tests*.

b- L'oreille

Les benchmarks fournis par les bases d'images d'oreilles IIT Delhi et USTB ont été utilisés en expérimentations. Deux scénarios expérimentaux ont été étudiés: le premier, utilise deux images référentielles en *apprentissage*, tandis que le deuxième, utilise uniquement une seule image référentielle en *apprentissage*.

La base d'images IIT Delhi [282] est constituée d'images d'oreilles, collectées par des étudiants et par le personnel de *l'université IIT Delhi*, New Delhi (Inde). Les images ont été acquises entre Octobre 2006 et Juin 2007 par l'utilisation d'une configuration simple de l'image. Toutes les images ont été prises à distance et dans un environnement intérieur. La base de données actuellement disponible a deux versions: la première version, dénotée par IIT Delhi-1, contient 493 images de 125 objets (personnes) différents et la deuxième, dénotée par IIT Delhi-2, contient 793 images de 221 objets différents. Chaque objet dans la base de

données comporte au moins trois images d'oreilles. Les objets sont dans l'intervalle d'âge de 14 à 58 ans. La résolution des images est de 272×204 pixels. La figure 5.4 représente certains échantillons. En addition aux images originales, cette base de données est également livrée avec des images d'oreilles automatiquement normalisées et recadrées dans une taille de 50×180 pixels (voir Figure.5.5).

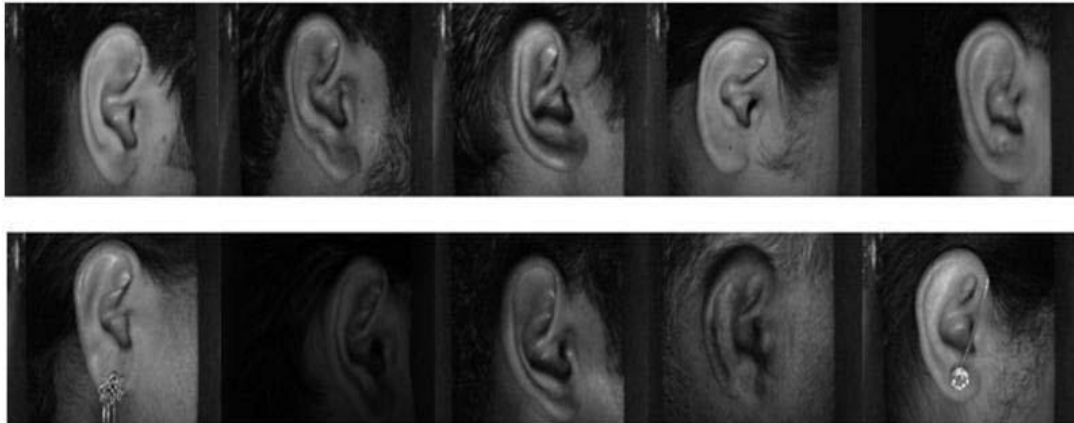


Figure 5. 4 : Quelques échantillons de la base d'images IIT Delhi.

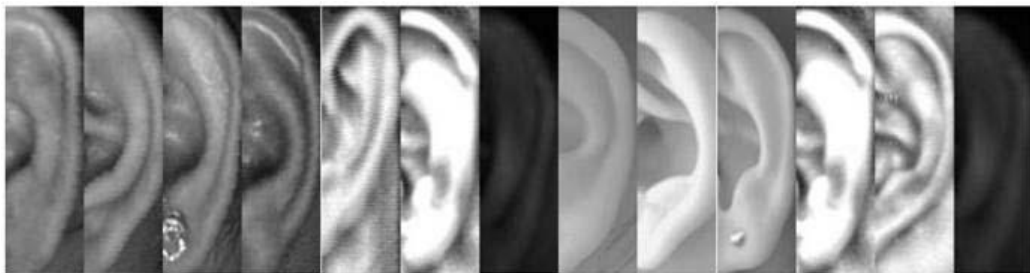


Figure 5. 5 : Quelques échantillons normalisés de la base d'images IIT Delhi

La base de données USTB [283] est constituée par des images d'oreilles collectées par *l'Université de la Science et de la Technologie de Beijing* (USTB). Les images de cette base de données ont été prises sous différentes conditions d'illumination et différents angles, elle contient 185 images de 60 objets au total. Il existe au moins trois images d'oreilles pour chaque personne. La base de données disponible fournit automatiquement des images d'oreilles normalisées et recadrées dans un cadre de 80×150 pixels. La figure 5.6 montre quelques échantillons de cette base d'images.

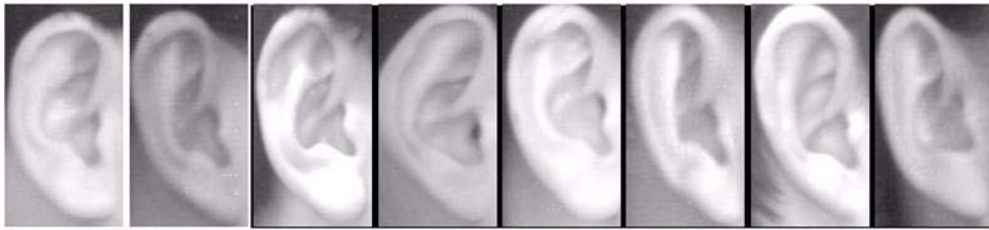


Figure 5. 6 : Quelques échantillons de la base d'images USTB.

Dans le premier scénario ("*expérimentations en utilisant deux images référentielles en apprentissage*"), deux images d'oreilles, pour chaque objet de la base de données, ont été utilisées comme un ensemble d'*apprentissage* et les images d'oreilles restantes (c'est-à-dire, entre une et quatre images) ont été utilisées comme un ensemble de *test*. Comme la plupart des objets dans les trois bases de données ont trois images, nous avons effectué trois permutations et rapporté le taux d'identification moyen au rang-1. Le nombre d'images dans les deux ensembles d'apprentissage et de test pour une permutation donnée, et pour chaque base d'images, est représenté dans le tableau 5.1.

Dans le deuxième scénario ("*expérimentations en utilisant une seule image référentielle en apprentissage*"), une seule image d'oreille, pour chaque personne dans la base de données, a été utilisée comme un ensemble d'*apprentissage*, tandis que les images d'oreilles restantes (c'est-à-dire, entre deux et quatre images) ont été utilisées comme un ensemble de *test*. Nous avons également fait trois permutations et rapporté le taux d'identification moyen au rang-1. Le nombre d'images dans les deux ensembles d'apprentissage et de test pour une permutation donnée, et pour chaque base d'images, est représenté dans le tableau 5.2.

Tableau 5. 1 : Distribution des images entre l'apprentissage et le test en utilisant deux images par personne en apprentissage.

Base d'images	IIT Delhi-1	IIT Delhi-2	USTB
Nombre total de personnes	125	221	60
Nombre total d'images	493	793	185
Images utilisées en apprentissage	250	442	120
Images utilisées en test	243	351	65

Tableau 5. 2 : Distribution des images entre l'apprentissage et le test en utilisant une seule image par personne en apprentissage.

Base d'images	IIT Delhi-1	IIT Delhi-2	USTB
Nombre total de personnes	125	221	60
Nombre total d'images	493	793	185
Images utilisées en apprentissage	125	221	60
Images utilisées en test	368	572	125

5.3.2 Expérimentation#1 (Effet du classificateur et le niveau de la décomposition)

Dans la première série d'expérimentations, nous avons utilisé les paramètres basiques de chaque descripteur de texture local; c'est-à-dire, l'opérateur 1DLBP a été pris avec $(m, W) = (9, 8)$, l'opérateur LPQ a été implémenté avec sa version basique ($rayon = 04$), l'opérateur BSIF a été implémenté avec un filtre de 11×11 et une longueur de chaîne binaire de 08 bits, tandis que l'opérateur LBP a été pris sous différents paramètres de (P, R) . Le choix de ces paramètres a été motivé par le fait que ces configurations, théoriquement et empiriquement, achèvent les meilleurs résultats.

Trois types de classificateurs ont été comparés et utilisés pour la classification, à savoir: k-NN, SVM avec un noyau *linéaire* et SVM avec un noyau *RBF*. Le classificateur k-NN a été implémenté avec la distance de type "*city block*". Dans le cadre de notre travail, nous avons trouvé, après plusieurs expérimentations, que cette distance présente les meilleurs résultats par rapport aux autres distances, telles que: "*Chi-Square*", "*Euclidienne*" ou "*Hamming*". La meilleure valeur du paramètre k pour le classificateur k-NN, trouvé, après plusieurs tests, est égale à 1. L'augmentation de la valeur du k affecte négativement les performances de la classification. Pour le SVM, nous avons choisi deux types de fonction noyau largement utilisés, à savoir: le noyau *linéaire* et le noyau *RBF*. Le paramètre γ de la fonction du noyau *RBF* a été choisi, dans le cadre de notre travail, d'une manière empirique ($\gamma = 0,0001$).

Les tableaux 5.3–5.7 comparent les résultats d'identification aux rang-1 en utilisant les trois classificateurs cités ci-dessus, les trois types de la projection horizontale-verticale, et

deux images par personne en apprentissage (le cas de l'oreille), appliqués sur les bases d'images ORL, AR, IIT Delhi-1, IIT Delhi-2 et USTB respectivement.

Les tableaux 5.8–5.10 comparent les résultats d'identification au rang-1, en utilisant une seule image par personne en apprentissage, les trois classificateurs cités ci-dessus et les trois types de la projection horizontale-verticale, appliqués sur les bases d'images IIT Delhi-1, IIT Delhi-2 et USTB respectivement.

Tableau 5.3 : Taux d'identifications au rang-1 avec différents classificateurs et niveaux de décomposition appliqués sur la base d'images ORL.

Classificateur	k-NN			SVM (noyau linéaire)			SVM (noyau RBF)		
	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3
LBP (8.1)	24.36	46.24	78.80	20.95	44	78.30	20.95	44	78.47
LBP (8.2)	26.74	48.44	79.70	21.35	45.74	79	21.35	45.74	78.64
LBP (8.3)	27.44	49	80.21	23.01	47.65	97.80	23.01	47.65	80
LBP (16.2)	26.01	46.24	79	22.14	44.98	78	22.14	44.98	78.14
1DLBP	30.14	51.47	86.60	30.14	49	86.10	30.14	49	86.64
LPQ basique	44.74	70.14	87.04	45	70.10	86.84	45	70.10	85.94
BSIF basique	86	88.34	92.14	85.47	87.95	91.92	85.47	88.50	92

Tableau 5.4 : Taux d'identifications au rang-1 avec différents classificateurs et niveaux de décomposition appliqués sur la base d'images AR.

Classificateur	k-NN			SVM (noyau linéaire)			SVM (noyau RBF)		
	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3
LBP (8.1)	15.24	32.85	64.14	14.96	30.48	61	14.96	30.48	61
LBP (8.2)	17.54	34.14	68.57	16.85	33.25	66.24	16.85	33.25	66.24
LBP (8.3)	18.22	34.65	69.14	17.02	34.65	68	17.02	34.65	68
LBP (16.2)	16.25	32.32	66.32	16.58	31.22	65.32	16.58	31.22	65.32
1DLBP	20.30	40.25	71.25	20.30	40.25	70.36	20.30	40.25	70.36
LPQ basique	36.54	55.14	76	34.65	52.58	76	34.65	52.58	76
BSIF basique	68.24	74.25	84.88	68	73.69	82.55	68	73.69	83.12

Tableau 5. 5 : Taux d'identifications au rang-1 en utilisant deux images par personne en apprentissage avec différents classificateurs et niveaux de décomposition appliqués sur la base d'images IIT Delhi-1.

Classificateur	k-NN			SVM (noyau linéaire)			SVM (noyau RBF)		
	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3
LBP (8.1)	38.46	38.63	62.82	27.43	51.17	77.5	25.1	51.17	73.94
LBP (8.2)	48.21	44.71	64.47	44.31	57.20	80.38	42.8	49.38	67.82
LBP (8.3)	40.08	45.40	66.39	40.33	62	81.62	34.43	51.03	62.55
LBP (16.2)	42.05	41.42	64.19	33.61	55.01	71.86	28.67	51.03	66.26
1DLBP	49,38	62.55	81.62	45.40	62	81.62	44.31	56.11	79.84
LPQ basique	58.97	70.78	81.8	59.4	70.23	79.84	60.08	59.55	79.97
BSIF basique	96.41	94.51	94.10	91.36	94.92	91.5	91.36	95.06	91.5

Tableau 5. 6 : Taux d'identifications au rang-1 en utilisant deux images par personne en apprentissage avec différents classificateurs et niveaux de décomposition appliqués sur la base d'images IIT Delhi-2.

Classificateur	k-NN			SVM (noyau linéaire)			SVM (noyau RBF)		
	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3
LBP (8.1)	20.41	35.23	59.82	24.12	46.63	74.74	21.65	45.96	71.51
LBP (8.2)	19.84	42.73	62.29	29.72	53.66	79.2	28.68	47.48	67.88
LBP (8.3)	23.64	42.07	64.67	36.47	59.07	80.53	32.48	50.24	64.96
LBP (16.2)	17.56	38.27	61.91	30.48	52.52	78.16	26.21	48.53	68
1DLBP	45.96	52.52	78.16	48.53	64.96	79.2	35.23	53.66	74.74
LPQ basique	55.55	70.08	81.95	57.27	67.62	79.77	57.83	66.86	79.87
BSIF basique	90.21	94.01	94.77	88.79	93.07	92.88	88.89	93.16	92.88

Tableau 5.7 : Taux d'identifications au rang-1 en utilisant deux images par personne en apprentissage avec différents classificateurs et niveaux de décomposition appliqués sur la base d'images USTB.

Classificateur	k-NN			SVM (noyau linéaire)			SVM (noyau RBF)		
	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3
LBP (8.1)	38.46	58.46	84.10	27.18	49.25	71.8	26.67	36.41	67.18
LBP (8.2)	48.21	68.71	90.26	37.95	64.1	85.64	38.46	53.85	69.74
LBP (8.3)	40.08	64.10	90.77	38.46	62.05	82.56	33.33	50.26	65.13
LBP (16.2)	42.05	66.66	88.71	40.51	62.05	83.59	34.87	52.82	69.23
1DLBP	49.25	69.23	90.26	41.54	62.05	83.59	40.08	53.85	71.28
LPQ basique	58.97	76.92	92.82	41.54	54.88	71.28	41.03	54.88	70.77
BSIF basique	96.41	98.46	97.94	89.23	94.36	97.44	89.23	94.36	97.44

Tableau 5.8 : Taux d'identifications au rang-1 en utilisant une seule image par personne en apprentissage avec différents classificateurs et niveaux de décomposition appliqués sur la base d'images IIT Delhi-1.

Classificateur	k-NN			SVM (noyau linéaire)			SVM (noyau RBF)		
	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3
LBP (8.1)	17.75	30.97	51.44	18.12	38.41	63.32	18.21	38.68	63.32
LBP (8.2)	19.83	36.41	53.98	26.72	43.48	67.12	26.45	43.66	67.12
LBP (8.3)	21.92	37.31	55.97	29.53	46.65	68.21	29.53	46.65	68.21
LBP (16.2)	18.93	34.32	53.07	23.61	41.58	67.03	24.18	41.67	67.03
1DLBP	29.53	43.48	67.12	26.72	41.58	63.32	26.45	41.67	67.03
LPQ basique	47.64	62.04	71.28	51.27	62.5	73.82	51.27	62.41	73.82
BSIF basique	83.42	88.42	86.95	85.33	88.5	85.78	85.33	88.5	85.78

Tableau 5.9 : Taux d'identifications au rang-1 en utilisant une seule image par personne en apprentissage avec différents classificateurs et niveaux de décomposition appliqués sur la base d'images IIT Delhi-2.

Classificateur	k-NN			SVM (noyau linéaire)			SVM (noyau RBF)		
	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3
LBP (8.1)	15.20	27.79	47.20	14.22	32.23	57.28	14.28	32.34	57.23
LBP (8.2)	16.02	33.04	50.23	21.21	37.76	62.94	20.98	37.82	62.94
LBP (8.3)	17.24	33.97	53.14	25.35	43.18	64.63	25.35	43.24	64.63
LBP (16.2)	14.91	30.53	49.47	19.87	37.3	63.4	17.76	37.35	63.46
IDLBP	24.32	36.74	65.20	24.32	33.12	63.4	23.69	34.52	61.94
LPQ basique	45.10	58.21	69.46	49.01	59.44	71.5	49.07	59.44	71.5
BSIF basique	79.95	85.78	86.65	85.9	85.49	81.47	85.9	85.49	81.47

Tableau 5.10 : Taux d'identifications au rang-1 en utilisant une seule image par personne en apprentissage avec différents classificateurs et niveaux de décomposition appliqués sur la base d'images USTB.

Classificateur	k-NN			SVM (noyau linéaire)			SVM (noyau RBF)		
	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3	Niveau 1	Niveau 2	Niveau 3
LBP (8.1)	28.8	45.07	73.87	20	31.73	55.2	20	31.73	54.93
LBP (8.2)	37.87	59.2	82.93	31.2	56.53	73.33	31.2	56.53	73.33
LBP (8.3)	34.4	55.47	81.07	28.8	49.07	69.6	28.8	49.07	69.6
LBP (16.2)	34.13	55.20	79.73	28.8	46.67	69.6	28.8	46.67	69.6
IDLBP	30.02	50.14	74.34	29.42	49.34	72.74	30.02	48.88	71.12
LPQ basique	44.80	59.3	81.60	40.53	52.27	66.93	40.53	52.27	66.93
BSIF basique	90.93	94.93	93.6	20	31.73	55.2	20	31.73	54.93

Comme il peut être remarqué, d'après une première analyse et interprétation des données des tableaux 5.3–5.10, nous constatons que les meilleurs résultats d'identifications appliquées sur les cinq bases de données et sous différents paramètres sont achevés, en général, avec le classificateur de type k-NN. Les résultats obtenus avec les deux classificateurs SVM-Linéaire et SVM-RBF sont presque similaires et sont moins performants que les résultats du k-NN.

Le descripteur LBP semble fonctionner mieux avec les banques de filtres à grande échelle que les banques de filtres à petite échelle. Ainsi, nous pouvons voir que le descripteur projeté sur un espace de 1D (1DLBP) est plus efficace que le descripteur bidimensionnel (LBP); en effet, les résultats obtenus avec le descripteur 1DLBP surpassent, en général, tous les résultats obtenus avec le descripteur LBP sous différentes configurations de (P, R) .

En outre, le niveau-3 de la décomposition horizontale-verticale est considéré comme la représentation qui semble de fournir les meilleurs taux d'identifications avec la plupart des descripteurs, à l'exception de BSIF où leurs résultats ne sont pas corrélés à un niveau spécifique.

Les résultats obtenus indiquent que l'opérateur LPQ surpasse, en général, le LBP et le 1DLBP sous les différentes configurations, tandis que le BSIF donne les meilleurs résultats avec toutes les bases d'images.

En fait, des résultats très intéressants par rapport à ce qui existe en état de l'art, sont apparus avec le descripteur BSIF, en utilisant uniquement une seule image d'oreille référentielle par personne en apprentissage.

Les résultats obtenus avec la base de données AR sont peu faibles par rapport à ceux obtenus avec les autres bases de données, ceci est expliqué par l'effet de l'occlusion; en fait, les images utilisées en apprentissage sont des images claires, simples et non-occluses, par contre les images employées en test sont des images masquées par des lunettes ou écharpes (images occluses). En outre, la décomposition horizontale-verticale a un effet très positif sur les performances de l'identification avec des images occluses; en fait, si un bloc est occlus, les autres blocs contiennent toujours un ensemble suffisant de caractéristiques distinctives pour compenser les informations perdues par l'occlusion.

Comme les meilleures performances sont achevées avec le BSIF en tant que extracteur de caractéristiques et le 1-NN comme classificateur, dans ce qui suit, nous nous baserons sur

cette configuration initiale et nous chercherons à améliorer et à optimiser les paramètres dans le but d'atteindre les meilleures performances d'identification.

5.3.3 Expérimentation#2 (Effet des paramètres de l'opérateur BSIF)

Afin de mieux comprendre les performances du descripteur BSIF, nous avons effectué plusieurs tests supplémentaires. Comme indiqué dans le chapitre précédent, le descripteur BSIF dépend de deux paramètres, qui sont: la taille de la fenêtre du filtre et le nombre de bits qui composent la chaîne du code binaire. Ainsi, pour trouver les paramètres optimaux qui donnent les meilleurs résultats, nous avons commencé par explorer tous les paramètres de l'opérateur BSIF proposés par *Kannala et Rahtu* (2012) [278]. Le tableau 5.11 récapitule les résultats d'application de tous les paramètres existants de l'opérateur BSIF; les paramètres ont été appliqués sur la base de données IIT Delhi-2 au niveau 3 de la décomposition horizontale-verticale, avec deux images par personne dans l'ensemble d'apprentissage et l'utilisation du 1-NN en tant que le meilleur classificateur trouvé dans l'expérimentation #1. Au lieu d'appliquer les tests supplémentaires sur toutes les bases de données, nous avons sélectionné uniquement la base d'images IIT Delhi-2 comme un échantillon, puisqu'elle est considérée comme la base de données la plus difficile et compliquée par rapport aux autres bases, les résultats obtenus avec toutes les bases de données sont presque similaires en terme d'analyse et d'interprétation de données, ainsi que l'application des tests supplémentaires sur toutes les bases de données qui nécessite trop de calculs et prend beaucoup de temps.

Tableau 5. 11 : Taux d'identifications au rang-1 en utilisant tous les paramètres existants de l'opérateur BSIF appliqués sur la base d'images IIT Delhi-2, niveau-3 de la projection horizontale-verticale, 1-NN classificateur et deux images par personne en apprentissage.

Paramètres BSIF	3×3	5×5	7×7	9×9	11×11	13×13	15×15	17×17
5 bits	72.43	78.6	84.77	85.19	85.6	87.24	87.24	85.19
6 bits	76.13	83.13	87.65	89.3	89.71	91.77	90.54	90.95
7 bits	74.9	84.36	89.3	90.54	94.65	95.06	92.59	92.18
8 bits	77.37	84.77	91.8	88.89	94.78	95.06	93.42	94.65
9 bits	/	83.13	86.83	90.12	95.06	92.59	95.88	95.47
10 bits	/	80.25	89.3	94.65	95.47	92.18	94.65	96.3
11 bits	/	82.3	91.36	94.65	95.88	94.65	96.3	96.3
12 bits	/	82.72	89.71	91.8	94.65	95.47	97.12	95.88

Il est clair d'après les résultats présentés dans le tableau 5.11 que le descripteur BSIF avec une banque de filtres de taille 15×15 pixels et une chaîne de caractères binaires à 12 bits surpasse toutes les combinaisons restantes. Les meilleurs paramètres de l'opérateur BSIF (filtre de 15×15 pixels et 12 bits) ont été appliqués, ensuite, sur les cinq bases d'images ORL, AR, IIT Delhi-1, IIT Delhi-2 et USTB. Les résultats obtenus, en utilisant le classificateur 1-NN et tous les types de la projection horizontale-verticale, sont présentés dans le tableau 5.12 et bien récapitulés dans la figure 5.7.

Tableau 5. 12 : Taux d'identifications au rang-1 en utilisant les paramètres optimaux de l'opérateur BSIF (filtre de taille 15×15 pixels et 12 bits).

Configuration	/		Deux images par personne en apprentissage			Une seule image par personne en apprentissage		
	ORL	AR	IIT Delhi-1	IIT Delhi-2	STB	IIT Delhi-1	IIT Delhi-2	STB
Niveau 1	89.51	73.12	97.26	97.34	98.46	94.29	92.6	96
Niveau 2	90.15	80.14	97.12	95.06	98.46	93.84	88.99	95.2
Niveau 3	94.36	87.67	96.61	97.12	97.2	91.21	90.25	94.93

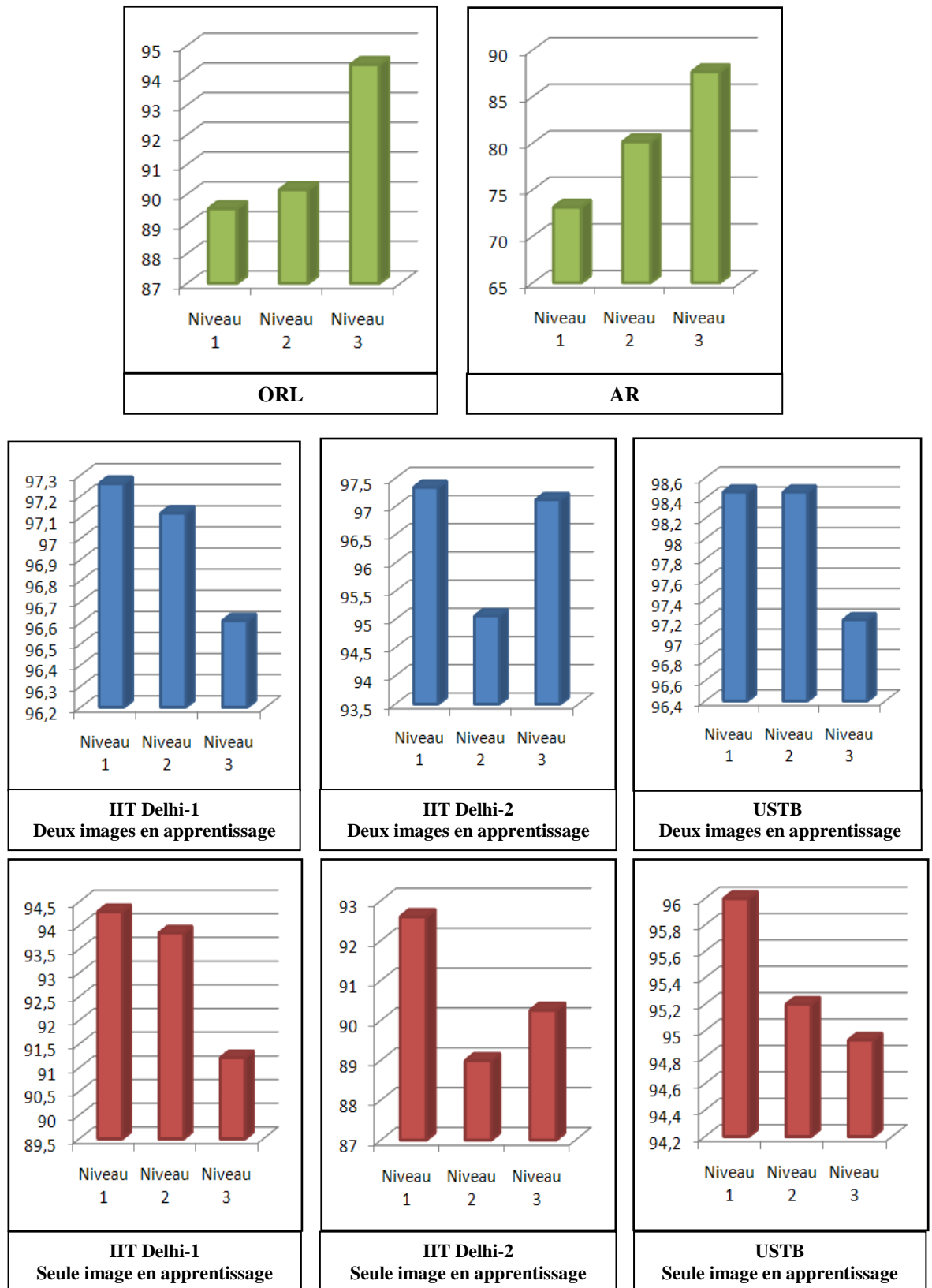


Figure 5.7 : Taux d'identifications au rang-1 en utilisant les paramètres optimaux de l'opérateur BSIF (filtre de taille 15×15 pixels et 12 bits).

Avec cette combinaison, nous avons particulièrement remarqué que c'est seulement avec l'utilisation du premier niveau de la décomposition horizontale-verticale, c'est-à-dire l'image entière, que nous pouvons atteindre de bonnes performances, en termes de taux d'identification et bien sûr en temps d'exécution, par rapport aux autres niveaux de décomposition. Ce résultat est valable uniquement pour le cas de l'oreille; les meilleurs résultats, pour le cas du visage, sont obtenus par la concaténation des caractéristiques extraites par les trois niveaux de décomposition. Ceci peut être expliqué par les démonstrations de la neuropsychologie en reconnaissance de visages (Chapitre 02, Section 2.3) qui suggèrent que les deux informations: globale (holistique) et locale (basée sur les caractéristiques) sont cruciales pour l'identification du visage (le niveau-3 de la décomposition horizontale-verticale englobe l'information globale (l'image entière) et l'information locale (les sous-blocs et les pixels)). Par ailleurs, dans le cas de l'oreille, l'information globale (c'est-à-dire, la forme de l'oreille) suffit uniquement à codifier l'information utile à l'identification des individus.

Contribution: D'après les résultats obtenus dans le cadre de notre travail, nous pouvons déduire que la reconnaissance du visage nécessite *vraiment* l'utilisation de composantes à basses et hautes fréquences qui contribuent à la description locale et globale respectivement (comme démontré en neuropsychologie). Par contre, la reconnaissance de l'oreille peut être accomplie avec succès en utilisant *uniquement* les composantes à hautes fréquences qui contribuent à la description de la forme globale de l'oreille.

5.3.4 Expérimentation #3 (Effet de l'information rigide / semi-rigide)

Sur la base des études neuropsychologiques du visage que nous avons effectué dans le Chapitre 02, Section 2.3, nous avons constaté que la région de la bouche est la plus sensible aux variations au cours des expressions faciales, tandis que la zone autour du nez est moins affectée par les expressions faciales en comparaison avec d'autres régions du visage. La figure 5.8 montre les variations d'une image portée de six expressions d'un individu, par rapport à une expression *neutre*. Comme nous pouvons le voir, les régions autour de la bouche ont des variations notables avec les expressions, telles que le *dégoût*, la *crainte* ou la *surprise*. Le sentiment de *surprise* conduit automatiquement à une réaction de bouche ouverte chez la plupart des personnes. Cette réaction est évidente même dans des images d'intensités. En outre, les régions autour des joues ont trop de variations au cours des expressions comme le *bonheur*. En revanche, les régions autour du nez ne sont pas altérées par les expressions

différentes pour la plupart des gens. Les régions des yeux et le front présentent la même situation.

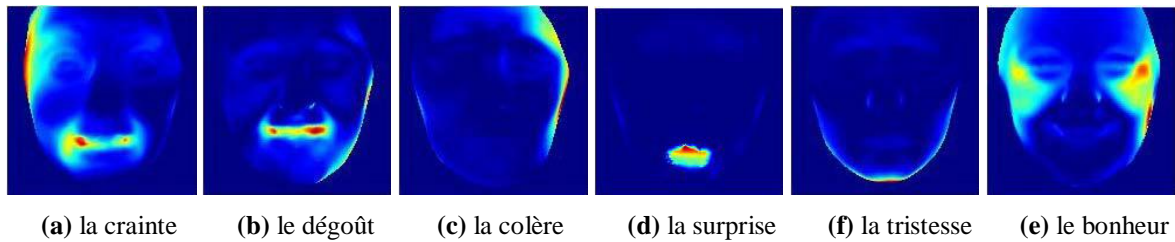


Figure 5.8 : Différence de la valeur de profondeur entre les balayages neutres et de six expressions pour une personne donnée.

Par conséquent, les régions rigides (autour du nez) et semi-rigides (bouche et menton) peuvent être séparées en utilisant un *masque* fournit par une technique heuristique simple. Pour chaque individu, le masque est défini manuellement pour inclure les yeux, le front et les régions du nez (information rigide) à l'aide d'une marge et d'exclure les régions de la bouche, les joues, le menton et les zones limites du visage (information semi-rigide). Avec ce masque, nous avons essayé de créer un équilibre sur les deux cotés de l'image du visage. Ce masque a été appliqué à chaque individu et pour chaque ensemble de données (apprentissage et test). La figure 5.9 illustre un exemple d'application du masque sur quelques échantillons de la base d'images AR. Son utilisation présente le grand avantage; en effet, elle élimine les calculs requis pour trouver les points de repères. En plus, le même masque a été utilisé pour toutes les personnes et il n'a pas été nécessaire de le régénérer à chaque fois.



Figure 5.9 : Exemple d'application du masque sur des échantillons de la base d'images AR, pour séparer entre les régions rigides et semi-rigides: (a) images par défaut et (b) images avec masque.

Sur la base des études anatomiques et embryologiques de l'oreille que nous avons effectué dans le Chapitre 03, Section 3.2, nous avons déterminé que la forme du lobule se change considérablement avec le vieillissement, en plus, le lobule se trouve fréquemment occlut ou encombré par des boucles. Nous n'avons pas prêté beaucoup d'attention à cette composante auriculaire non stable et nous l'avons considéré comme une région semi-rigide, les autres régions ont été considérées rigides. La figure 5.10 montre un exemple de séparation entre les régions rigides et semi-rigides.

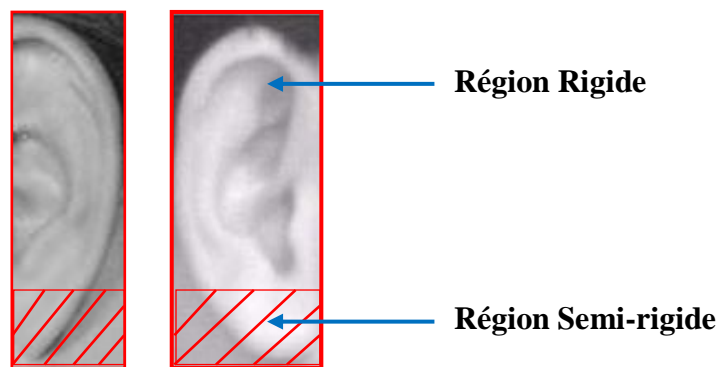


Figure 5. 10 : Exemple de séparation entre les régions rigides et semi-rigides.

Sur la base des études que nous avons effectué sur le descripteur BSIF dans le Chapitre 04, Section 4.7, nous avons remarqué que parmi les limitations de l'opérateur BSIF, les filtres originaux de ce descripteur ont été entraînés à partir d'un ensemble naturel et aléatoire composé de 50 000 patches d'images texturées (information semi-rigide). Pour être plus précis, nous avons proposé, dans le cadre de cette thèse, de refaire l'apprentissage des filtres de ce descripteur, mais, uniquement par des images de visages et d'oreilles respectivement (information rigide), et non pas avec un ensemble d'images naturelles, aléatoires, et texturées (le nouvel apprentissage des filtres est bien détaillé dans le Chapitre 04, Section 4.7).

Les performances de l'identification au rang-1 en utilisant les informations rigides et semi-rigides pour le visage, l'oreille et l'opérateur BSIF et bien sûr avec les meilleures configurations déterminées dans les expérimentations #1 et #2 sont récapitulées dans le tableau 5.13. Nous pouvons y voir que les nouveaux pourcentages de la classification obtenus par l'utilisation des informations rigides ont été améliorés d'une manière très significative,

spécialement avec la base d'images AR où la plupart des régions sensibles aux expressions et contenant des occlusions de l'écharpe sont presque éliminées par le masque.

Tableau 5. 13 :Taux d'identifications au rang-1 en utilisant les informations rigides et semi-rigides appliqués sur les différentes bases d'images.

Configuration	/		Deux images par personne en apprentissage			Une seule image par personne en apprentissage		
	ORL	AR	IIT Delhi-1	IIT Delhi-2	USTB	IIT Delhi-1	IIT Delhi-2	USTB
Informations Semi-rigides	94.36	87.67	97.26	97.34	98.46	94.29	92.6	96
Informations Rigides	97	93.36	98	97.49	98.46	95.34	94.22	97.62

Afin de mieux comprendre l'effet des expressions et de l'occlusion sur la base d'images AR, nous avons utilisé un nouveau protocole d'évaluation B lié à cette base de données, il consiste à utiliser les images du premier type (*expression neutre*) en apprentissage, et tous les autres types (*sourire, colère, cri perçant, lunette et écharpe (voir figure.5.3)*) sont utilisés séparément comme des ensembles de test. Les résultats d'identification de ce protocole sont récapitulés dans le tableau 5.14.

Tableau 5. 14 : Taux d'identifications au rang-1 en utilisant le "Protocole B" appliqué sur la base d'images AR.

Expression	Sourire	Colère	Cri perçant	Avec Lunette	Avec Echarpe
Taux d'identification au rang-1 %	100	100	93.24	90.84	97.26

Comme interprétation de ces résultats, nous pouvons observer des résultats très puissants avec les ensembles de test type "Sourire" et "Colère", les résultats d'identification avec l'ensemble "Cri Perçant" sont peu faibles puisque la forme des yeux humains et du nez se changent considérablement avec ce type d'expression, tandis que les taux d'identification avec l'ensemble "Avec Lunettes" sont faibles par rapport à l'ensemble "Avec Echarpe", parce que les yeux humains, représentant les caractéristiques faciaux les plus importants en identification de visages, sont cachés.

5.3.5 Comparaison

Pour une analyse complète, nous avons également comparé les résultats obtenus avec ceux de l'état de l'art pour l'identification automatique du visage / oreille. Les deux tableaux 5.15 et 5.16 montrent et comparent le taux d'identification moyen au rang-1 de la méthode proposée avec quelques approches d'extraction des caractéristiques récentes et bien connues, dans les mêmes conditions avec le même protocole d'évaluation (Pour la base d'images AR, nous avons employé le protocole "B", tandis que le protocole le plus utilisé dans la littérature, pour le cas de l'oreille, est de type: "*deux images par personne en apprentissage*").

Tableau 5. 15 : Sommaire des travaux connexes et récents en identification des individus par visages.

Référence	Extracteur de caractéristiques	ORL	AR				
			Sourire	Colère	Cri perçant	Avec Lunette	Avec Echarpe
Oh <i>et al.</i> (2008) [284]	S-LNMF	-	62	-	27	49	55
Min <i>et al.</i> (2011) [285]	/	-	-	-	52.3	54.2	81.3
Nguyen <i>et al.</i> (2012) [286]	ELBP		95	98.3	57.2	57.2	95
Nguyen and Caplier (2012) [287]	ELBP + WPCA	-	96.6	98.3	62.2	65.6	95.8
Luan <i>et al.</i> (2014) [288]	RPCA	-	-	-	-	90.5	89.5
Zhao <i>et al.</i> (2014) [289]	PCA	92.4	-	-	-	-	-
	RBF-PCA	93.2	-	-	-	-	-
	SK-PCA	95.25	-	-	-	-	-
Li <i>et al.</i> (2014) [290]	RCBD	96.7	-	-	-	93.9	92.8
Notre Approche	BSIF	97	100	100	93.24	90.84	97.26

Tableau 5. 16 : Sommaire des travaux connexes et récents en identification des individus par oreilles.

Référence	Extracteur de caractéristiques	Delhi-1	Delhi-2	USTB
Hurley et al. (2005) [179,184]	Force Field Transform	74.93	66.67	/
Anam and Usman (2012) [291]	Haar wavelets	95.2	/	98.33
Kumer and Wu (2012) [184]	Shape feature	29.07	30.62	/
	Gabor phase	83.47	84.46	/
	Gabor orientation weighted	90.4	88.39	/
	Orthogonal log-Gabor filter Pair	96.27	95.93	/
Chan and Kumer (2012) [292]	1D quadrature filter	95.73	94.72	/
	2D quadrature filter	96.53	96.08	/
Kumer and Chan (2013) [293]	Sparse representation of local gray-level orientations	97.07	97.73	/
Mamta and Madasu (2013) [294]	Traditional PCA	82.6	79	/
	Local Principal Independent Components	97.6	97.2	/
Basit et al. (2014) [191]	Non linear curvelet features	97.77	96,22	/
Jacob and Raju (2014) [295]	Gray Level Co-occurrence Matrix (GLCM) + LBP(16,2) + Gabor Filter	/	94.12	/
Notre Approche	BSIF	98	97.49	98.46

A partir de ces deux tableaux, nous pouvons dire que l'approche d'identification biométrique que nous proposons, soit pour le visage ou l'oreille, basée principalement sur le descripteur de texture local BSIF, montre des performances très compétitives surpassant la plupart des approches récentes de l'état de l'art, dans les mêmes conditions.

Il est également intéressant, dans des futurs travaux, d'évaluer notre approche proposée sur des bases de données à grande échelle impliquant des millions d'identités. Il est aussi intéressant d'évaluer l'approche proposée dans des contextes de vidéos surveillances qui exigent des mécanismes de détection contre les attaques de type: "*Spoofing Attacks*" et une identification en temps réel avec des vidéos à faible résolution.

5.4 Conclusion

Dans ce chapitre, nous avons comparé et testé les performances de quatre descripteurs de texture locaux très récents dans le domaine de l'analyse de texture, à savoir LBP, 1DLBP, LPQ et BSIF, appliqués sur des données biométriques bidimensionnelles, en utilisant des bases d'images de visages et/ou oreilles en expérimentations. Les résultats obtenus montrent, de manière générale, que le descripteur BSIF permet de mieux différencier et identifier les individus. Des efforts additionnels ont été fournis pour améliorer les taux d'identifications, tels que: la recherche des paramètres optimaux du descripteur BSIF, en termes de taille du filtre et de la chaîne de caractères binaires, ainsi que l'utilisation des informations rigides/semi-rigides pour le visage, l'oreille et le BSIF. En conclusion, l'application de l'approche que nous proposons, basée sur le BSIF, sur des bases de données à grande échelle, dans des environnements de vidéo surveillance, et pour un système biométrique multimodal impliquant le visage et l'oreille simultanément, est nécessaire, dans des futurs travaux, afin de mieux mesurer et comprendre ses performances.

Conclusion et Perspectives

Les systèmes biométriques sont devenus des outils de plus en plus importants pour la sécurité de l'individu et de l'information, dans les sociétés modernes largement interconnectées; ils fournissent une vérification ou une identification automatique de l'identité basée sur une analyse biologique de l'individu. En littérature, plusieurs modalités ont été étudiées et comparées. Nous nous sommes intéressés dans cette thèse par les modalités faciales, en particulier le visage et l'oreille, puisqu'elles fournissent une identification naturelle, non-intrusive et très acceptable par le public.

Les images de données biométriques bidimensionnelles peuvent être considérées comme une combinaison de caractéristiques globales et locales, qui peuvent être bien décrites à l'aide des descripteurs de texture locaux. Dans le cadre de ce travail, nous avons exploité cette observation, qui imite exactement les capacités du système neurologique humain dans la perception et la reconnaissance des objets et des visages, et nous avons proposé une représentation simple et efficace pour l'identification biométrique des individus, en particulier, par l'utilisation de l'apparence de leurs visages ou oreilles. La représentation proposée peut être facilement extraite dans un balayage simple de l'image et sans aucune analyse complexe.

Pour un fonctionnement efficace de l'approche proposée, nous avons introduit, tout d'abord, un prétraitement préalable, qui consiste à préparer la représentation de l'image originale afin de faciliter la tâche des modules suivants et d'améliorer les performances de l'identification, en appliquant le filtre *médian* et la normalisation de l'histogramme.

L'image normalisée a été décomposée en plusieurs sous-images de différentes tailles en utilisant la projection horizontale-verticale, afin de mieux caractériser les détails et la corrélation entre les patches de l'image. En effet, plusieurs niveaux de décomposition ont été testés et comparés.

Pour la caractérisation de données, quatre descripteurs de texture locaux, qui sont: LBP, 1DLBP, LPQ et BSIF, ont été appliqués indépendamment sur chaque bloc (sous-image) et comparés, et par conséquent, un histogramme a été extrait à partir de chaque bloc. Les

histogrammes extraits de chaque bloc ont été concaténés dans un histogramme de caractéristiques global, pour chaque descripteur, représentant l'image du visage ou l'oreille.

Pour le visage, des études approfondies en psychologies et en neurosciences ont été effectuées dans le but d'identifier l'ensemble de composantes autonomes et aussi leurs emplacements sur lesquels nous pouvons atteindre une grande variation interindividuelle. Tandis qu'avec l'oreille, des études sur l'embryologie et les malformations de la forme extérieure de l'oreille ont été faites afin de mieux comprendre sa stabilité et son développement au cours du temps. Ces études sont particulièrement bénéfiques à cette recherche puisqu'elles fournissent également des conseils sur les variations possibles de la structure extérieure de chaque modalité.

En conséquence de ces études, nous avons proposé, dans le cadre de cette thèse, l'utilisation des masques pour séparer entre les régions *rigides* et *semi-rigides* de chaque modalité. Par exemple (le cas du visage), un masque a été implémenté manuellement pour *exclure* les régions de la bouche, les joues, le menton et les cotés frontières du visage, qui sont considérés comme des régions non-stables (région semi-rigide), et d'*inclure* les yeux, le front et les régions du nez qui sont considérées comme des informations importantes pour la reconnaissance (région rigide).

Une fois la représentation est établie, les trois classificateurs: les k-plus proches voisins (k-NN), les séparateurs à vaste marge (SVM) avec un noyau *linéaire* et les séparateurs à vaste marge (SVM) avec un noyau *RBF* ont été comparés et utilisés pour la classification.

Plusieurs tests ont été réalisés sur des images bidimensionnelles de quatre bases de données (ORL, AR, IIT Delhi et USTB) afin de comparer les performances de notre approche avec les quatre descripteurs de texture locaux discutés précédemment. Deux bases d'images comportant des images de visages tandis que les deux autres comportant des images d'oreilles, acquises chez différentes personnes et sous différentes conditions; en effet, chaque base de données a été décomposée en deux ensembles, le premier a été utilisé pour l'*apprentissage* tandis que le deuxième a été employé pour le *test*, en plus chaque base de données a son propre protocole d'évaluation. Ainsi, les effets de chaque classificateur, le niveau de la décomposition horizontale-verticale, les paramètres de chaque descripteur ainsi que l'utilisation de l'information rigide/semi-rigide ont été testés et les meilleurs paramètres ont été retenus.

Les résultats obtenus ont montré, d'une manière générale, que notre approche proposée a achevé des résultats très compétitifs par rapport à l'état de l'art. En plus, le descripteur BSIF permet de mieux différencier et identifier les individus en comparaison avec les autres descripteurs considérés. Les meilleurs taux d'identification ont été obtenus avec la configuration suivante: le classificateur 1-NN avec la distance "*city block*", le niveau-03 de la décomposition pour le visage, le niveau-01 de la décomposition pour l'oreille, l'utilisation de l'information rigide pour chaque modalité ainsi que l'utilisation d'un nouvel apprentissage spécialisé pour les filtres W du descripteur BSIF.

Enfin, les différents travaux de recherche menés dans le cadre de cette thèse ouvrent différentes perspectives pour des futurs travaux:

- Il serait intéressant de comparer d'autres classificateurs récents, tels que: les forêts aléatoires (RF: Random Forests), les régresseurs par vecteurs de supports (SVR: Support Vector Regressors) ou les techniques de projection comme: l'analyse de la corrélation canonique (CCA: Canonical Correlation Analysis), pour mieux discriminer les individus. En plus, il serait intéressant de tester l'approche proposée dans des conditions d'éclairage dégradées, dans un contexte de vidéo surveillance à faible résolution et sur des bases d'images à grande échelle impliquant des millions d'identités, afin de consolider les résultats obtenus.
- Le contrôle d'accès basé sur l'identification faciale (visage ou oreille) est seulement un exemple d'une application d'analyse faciale. Notre objectif principal consiste à développer l'approche proposée vers un système complet qui peut efficacement détecter, faire le suivi dans une vidéo, reconnaître, analyser les expressions faciales, animer les attributs faciaux et pourquoi pas imiter la plupart des capacités remarquables du système visuel humain. Quoique nous ne soyons pas là encore, les progrès courants dans le domaine nous donnent beaucoup de raisons d'être optimistes.
- Il serait aussi intéressant de faire la fusion multimodale du visage et l'oreille simultanément, dans le but d'établir l'identité d'un individu avec une précision très élevée. La fusion d'informations peut être accomplie à différents niveaux dans le système biométrique y compris le niveau capteur, le niveau caractéristique ou le niveau décision.

- Nous pensons que l'approche proposée dans le cadre de ce travail peut être également appliquée à plusieurs autres tâches de détection ou reconnaissance. Pour l'instant, nous voudrions exploiter le domaine de la *biométrie cachée* nommé aussi *la biométrie médicale*, qui constitue le futur de la biométrie en tant que nouvel axe de recherche émergeant dans la criminalistique et la sécurité des individus.

Mot de la fin

"Les travaux effectués durant ces trois années de thèse m'ont permis, d'une part d'approfondir mes connaissances en biométrie et d'autre part, d'y apporter une modeste contribution à la recherche scientifique, notamment dans les deux contextes de la sécurité des individus et de la criminalistique."

Liste des Publications, Communications Indexées, et Evènements Scientifiques

- A. Benzaoui, A. Hadid, and A. Boukrouche: "*Ear Biometric Recognition using Local Texture Descriptors*". Journal of Electronic Imaging (JEI-SPIE). Vol.23, No.05, pp.053008, 2014.
- A. Benzaoui and A. Boukrouche: "*Face Analysis, Description, and Recognition using Improved Local Binary Patterns in One Dimensional Space*". Journal of Control Engineering and Applied Informatics (CEAI). Vol.16, No.04, pp.52-60, 2014.
- A. Benzaoui, N. Hezil, and A. Boukrouche: "*Ear Recognition using Elliptical Local Binary Patterns*". In the 2nd International Conference on Signal, Image, Vision and their Applications (SIVA). Pp.117-121, November 18-20, 2013, Guelma (Algeria).
- A. Benzaoui: was attended the "**10th Summer School for Advanced Studies on Biometrics for Secure Authentication: HOW BIOMETRICS MEETS FORENSICS, SECURITY AND THE E-SOCIETY FOR TOMOROW**". June 10-15, 2013, Alghero (Italy).
- A. Benzaoui and A. Boukrouche: "*Face Recognition using 1DLBP Texture Analysis*". In the 5th International IARIA Conference on Future Computational Technologies and Applications (Future Computing). Pp.14-19, May 27 – June 01, 2013, Valencia (Spain).
- A. Benzaoui and A. Boukrouche: "*1DLBP and PCA for Face Recognition*". In the 11th International IEEE Symposium on Programming and Systems (ISPS). Pp.07-11, April 22-24, 2013, Algiers (Algeria). (**3rd Best Paper Award**)
- A. Benzaoui, H. Bourouba, and A. Boukrouche: "*System for Automatic Faces Detection*". In the 3rd International IEEE Conference on Image Processing, Theory, Tools and Applications (IPTA). Pp.354-358, October 15-18, 2012, Istanbul (Turkey).
- A. Benzaoui and H.F. Merouani: "*Hybrid System for Robust Faces Detection*". Journal of Electronic Science and Technology (JEST). Vol.10, No.02, pp.167-172, 2012. (**Special Issue on Signal and Image Processing**)

Références Bibliographiques

- [1] J.N. Pato and L.I. Millet, editors: *Biometric Recognition: Challenges and Opportunities*. D.C, The National Academic Press, Washington (USA), 2010.
- [2] S.Z. Li and A.K. Jain, editors: *Encyclopedia of Biometrics*. Springer Science + Business Media, New York (USA), 2009.
- [3] J.L. Wayman, A.K. Jain, D. Maltoni, and D. Maio, editors: *Biometric Systems: Technology, Design and Performance Evaluation*. Springer Verlage, London (UK), 2005.
- [4] L. O’Gorman: *Comparing Passwords, Tokens, and Biometrics for User Authentication*. Proceedings of the IEEE. Vol.91, No.12, pp.2021-2040, 2003.
- [5] A.K. Jain, A.A. Ross, and K. Nandakumar, editors: *Introduction to Biometrics*. Springer Science + Business Media, New York (USA), 2011.
- [6] M.D. Femila and A.A. Irudhayaraj: *Biometric System*. In Proceedings of the 3rd IEEE International Conference on Electronics Computer Technology (ICECT). pp.152-156, Kanyakumari (India), 2011.
- [7] K. Delac and M. Grgic: *A Survey of Biometric Recognition Methods*. In Proceedings of the 46th IEEE International Symposium of Electronics in Marine (ELMAR). pp.184-193, Zadar (Croatia), 2004.
- [8] D.O. Gorodnichy: *Evolution and Evaluation of Biometric Systems*. In Proceedings of the IEEE Symposium on Computational Intelligence in Security and Defence Applications (CISDA). pp.1-8, Ottawa (Canada), 2009.
- [9] S. Prabhakar, S. Pankanti, and A.K. Jain: *Biometric Recognition: Security and Privacy Concerns*. IEEE Security & Privacy. Vol.01, No.02, pp.33-42, 2003.
- [10] A.Y.J. Nakanishi and B.J. Western: *Advancing the State-of-the-Art in Transportation Security Identification and Verification Technologies: Biometric and Multibiometric Systems*. In Proceedings of the IEEE Intelligent Transportation Systems Conference. pp.1004-1009, Seattle (USA), 2007.
- [11] P.J. Phillips, A. Martin, C.L. Wilson, and M. Przybocki: *An Introduction to Evaluating Biometric Systems*. IEEE Computer. Vol.33, No.02, pp.56-63, 2000.
- [12] M. Golfarelli, D. Maio, and D. Maltoni: *On the Error-Reject Trade-Off in Biometric Verification Systems*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.19, No.07, pp.786-796, 1997.

- [13] S. Prabhakar and A.K. Jain: *Decision-Level Fusion in Fingerprint Verification*. Pattern Recognition (Elsevier). Vol.35, No.04, pp.861-874, 2002.
- [14] T. Fawcett: *An Introduction to ROC Analysis*. Pattern Recognition Letters (Elsevier). Vol.27, No.08, pp.861-874, 2006.
- [15] R.M. Bolle, J.H. Connell, S. Pankanti, N.K. Ratha, and A.W. Senior: *The Relation between the ROC Curve and the CMC*. In Proceedings of the 4th IEEE Workshop on Automatic Identification Advanced Technologies. pp.15-20, New York (USA), 2005.
- [16] A.K. Jain, R. Bolle, and S. Pankati, editors: *Biometrics: Personal Identification in Network Society*. Springer-Verlag, New York (USA), 1999.
- [17] A. Benzaoui and A. Boukrouche: *1DLBP and PCA for Face Recognition*. In Proceedings of the 11th IEEE International Symposium on Programming and Systems (ISPS). pp.7-11, Algiers (Algeria), 2013.
- [18] M.P. Beham and S.M.M Roomi: *A Reiew of Face Recognition Methods*. International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI). Vol.27, No.04, pp.1-35, 2013.
- [19] J. Daugman: *How Iris Recognition Works ?*. IEEE Transactions on Circuits and Systems for Video Technology. Vol.14, No.01, pp.21-30, 2004.
- [20] J. Daugman: *Recognizing People by their Iris Patterns*. Information Security Technical Report (Elsevier). Vol.03, No.01, pp.33-39, 1998.
- [21] C. Fancourt, L. Bogoni, K. Hanna, Y. Guo, R. Wildes, N. Takahachi, and U. Jain: *Iris Recognition at a Distance*. In Proceedings of the 5th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA). Lecture Notes in Computer Science (Springer). Vol.3546, pp.1-13, New York (USA), 2005.
- [22] F. Ahmed and D. Mohamed: *A review on fingerprint classification techniques*. In Proceedings of the International IEEE Conference on Computer Technology and Development (ICCTD). Vol.2, pp.411-415, Kota Kinabalu (Malaisie), 2009.
- [23] D. Maltoni, D. Maio, A.K. Jain, and S. Prabhakar, editors: *Handbook of Fingerprint Recognition (2nd Edition)*. Springer Verlag, 2009.
- [24] P. Varchol and D. Levicky: *Using of Hand Geometry in Biometric Security Systems*. Radioengineering. Vol.16, No.04, pp.82-87, 2007.
- [25] N. Saxena, V. Saxena, N. Dubey, and P. Mishra: *Hand Geometry: A New Method for Biometric Recognition*. International Journal of Soft Computing and Engineering (IJSCE). Vol.02, No.06, pp.192-196, 2013.

- [26] D. Zhang, Z. Guo, G. Lu, D. Zhang, and W. Zuo: *An Online System of Multispectral Palmprint Verification*. IEEE Transactions on Instrumentation and Measurement. Vol.59, No.02, pp.480-490, 2010.
- [27] Z. Guo, D. Zhang, L. Zhang, W. Zuo, and G. Lu: *Empirical Study of Light Source Selection for Palmprint Recognition*. Pattern Recognition Letters (Elsevier). Vol.32, No.02, pp.120-126, 2011.
- [28] B. Arbab-Zavar and M.S. Nixon: *On Guided Model-Based Analysis for Ear Biometrics*. Computer Vision and Image Understanding (Elsevier). Vol.115, No.04, pp.487-502, 2011.
- [29] A. Ross and A. Abaza: *Human Ear Recognition*. IEEE Computer, Biometrics Compendium. Vol.44, No.11, pp.79-81, 2011.
- [30] Z. Huang, Y. Liu, C. Li, M. Yang, and L. Chen: *A Robust Face and Ear based Multimodal Biometric System using Sparse Representation*. Pattern Recognition (Elsevier). Vol.46, No.08, pp.2156-2168, 2013.
- [31] J.P. Campbell: *Speaker Recognition: A Tutorial*. Proceedings of the IEEE. Vol.85, No.09, pp.1437-1462, 1997.
- [32] F. Monroe and A.D. Rubin: *Keystroke Dynamics as a Biometric for Authentication*. Future Generation Computer Systems (Elsevier). Vol.16, No.04, pp.351-359, 2000.
- [33] A.K. Jain, F.D. Griess, and S.D. Connell: *On-Line Signature Verification*. Pattern Recognition (Elsevier). Vol.35, No.12, pp.2963-2972, 2002.
- [34] I. Bouchrika: *Gait Analysis and Recognition for Automated Visual Surveillance*. Doctoral Thesis in Philosophy. School of Electronics and Computer Science, Faculty of Engineering, Science and Mathematics. University of Southampton (UK), 2008.
- [35] Y. Wang, Y. Fan, W. Liao, K. Li, L. Shark, and M.R. Varely: *Hand Vein Recognition based on Multiple Keypoints Sets*. In Proceedings of the 5th IEEE International Conference on Biometrics (ICB). pp.367-371, New Delhi (India), 2012.
- [36] Y.P. Hu, Z.Y. Wang, X.P. Yang, and Y.M. Xue: *Hand Vein Recognition based on the Connection Lines of Reference Point and Feature Point*. Infrared Physics & Technology (Elsevier). Vol.62, pp.110-114, 2014.
- [37] W. Song, T. Kim, H.C. Kim, J.H. Choi, H.J. Kong, and S.R. Lee: *A Finger-Vein Verification System using Mean Curvature*. Pattern Recognition Letters (Elsevier). Vol.32, No.11, pp.1541-1547, 2011.
- [38] L. Yang, G. Yang, Y. Yin, and X. Xi: *Exploring Soft Biometric Trait with Finger Vein Recognition*. Neurocomputing (Elsevier). Vol.135, pp.218-288, 2014.
- [39] S. Liu and M. Silverman: *A Practical Guide to Biometric Security Technology*. IEEE IT Professional. Vol.03, No.01, pp.27-32, 2001.

- [40] A.A. Ross, K. Nandakumar, and A.K. Jain: *Handbook of Multibiometrics*. Springer Science + Business Media (International Series on Biometrics). Vol.06, New York (USA), 2006.
- [41] A. Benzaoui and A. Boukrouche: *Face Recognition using 1DLBP Texture Analysis*. In Proceedings of the 5th International Conference on Computational Technologies and Applications (IARIA). pp.14-19, Valencia (Spain), 2013.
- [42] A. Hadid: *Face Biometrics under Spoofing Attacks: Vulnerabilities, Countermeasures, Open Issues, and Research Directions*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp.113-118, Ohio (USA), 2014.
- [43] A. Nait-Ali: *Hidden Biometrics: Towards using Biosignals and Biomedical Images for Security Applications*. In Proceedings of the 7th IEEE Workshop on Systems, Signal Processing and their Applications (WOOSPA). pp.352-356, Tipaza (Algeria), 2011.
- [44] K. Aloui, A. Nait-Ali, and M.S. Naceur: *New Biometric Approach based on Geometrical Human Brain Patterns Recognition: Some Preliminary Results*. In Proceedings of the 3rd IEEE European Workshop on Visual Information Processing (EUVIP). pp.258-263, Paris (France), 2011.
- [45] K. Aloui, A. Nait-Ali, and M.S. Naceur: *A Novel Approach based on Brain Biometrics: Some Preliminary Results for Individual Identification*. In Proceedings of the IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM). pp.91-95, Paris (France), 2011.
- [46] K. Aloui, A. Nait-Ali, and M.S. Naceur: *A New Useful Biometrics Tool based on 3D Brain Human Geometrical Characterizations*. Journal of Signal and Information Processing (JSIP). Vol.03, No.02, pp.198-207, 2012.
- [47] A. Nait-Ali: *Beyond Classical Biometrics: When using Hidden Biometrics to Identify Individuals*. In Proceedings of the 3rd IEEE European Workshop on Visual Information Processing (EUVIP). pp.241-246, Paris (France), 2011.
- [48] S.A. Hameed, B.B. Zaidan, A.A. Zaidan, A.W. Naji, and O.F. Tawfiq: *An Accurate Method to Obtain Bio-Metric Measurements for Three Dimensional Skull*. Journal of Applied Sciences. Vol.10, No.02, pp.145-150, 2010.
- [49] H.O. Alanazi, B.B. Zaidan, and A.A. Zaidan: *3D Skull Recognition using 3D Matching Technique*. Journal of Computing. Vol.02, No.01, 121-126, 2010.
- [50] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell: *Face recognition by humans: Nineteen results all computer vision researchers should know about*. Proceedings of the IEEE. Vol.94, No.11, pp.1948-1962, 2006.

- [51] G. Porter and G. Doran: *An anatomical and photographic technique for forensic facial identification*. Forensic Science International (Elsevier). Vol.114, No.2, pp.97-105, 2000.
- [52] V. Bruce, editor: *Recognizing faces*. Lawrence Erlbaum Associates Ltd, London (UK), 1988.
- [53] J.W. Shepherd, G.M.Davies, and H.D. Ellis: *Studies of cue saliency*. In G.M. Davies, H.D. Ellis, and J.W. Shepherd, editors: *Perceiving and remembering faces*. Academic Press, London (UK), pp.105-132, 1981.
- [54] K. Etemad and R. Chellapra: *Discriminant analysis for recognition of human face images*. Journal of the Optical Society of America A. Vol.14, No.8, pp.1724-1733, 1997.
- [55] P.J.B. Hancock, V. Bruce, and M. Burton: *A comparison of two computer-based face identification systems with human perceptions of faces*. Vision Research (Elsevier). Vol.38, No.15-16, pp.2277-2288, 1998.
- [56] P. Kalocsai, W. Zhao, and E. Elagin: *Face similarity space as perceived by humans and artificial systems*. In proceeding of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition (AFGR). pp.177-180, Nara (Japan), 1998.
- [57] A. Johnston, H. Hill, and N. Carman: *Recognizing faces: Effects of lighting direction, inversion, and brightness reversal*. Perception. Vol.21, No.3, pp.365-375, 1992.
- [58] I. Biederman and P. Kalocsai: *Neural and psychophysical analysis of object and face recognition*. In H. Wechsler, P.J. Phillips, V. Bruce, F.F. Soulié, and T.S. Huang, editors: *Face recognition: From theory to applications*. Vol.163, Part.1, Springer-Verlag, Berlin (Germany), pp.3-25, 1998.
- [59] H.D. Ellis: *Introduction to aspects of face processing: Ten questions in need of answers*. In H.D. Ellis, M.A. Jeeves, F. Newcombe, and A. Young, editors: *Aspects of face processing*. Vol.28, Part.1, Springer Netherlands, pp.3-13, 1986.
- [60] I. Gauthier, M. Behrmann, and M.J. Tarr: *Can face recognition really be dissociated from object recognition ?*. Journal of Cognitive Neuroscience (MIT Press). Vol.11, No.4, pp.349-370, 1999.
- [61] I. Gauthier and N.K. Logothetis: *Is face recognition not so unique after all ?*. Cognitive Neuropsychology (Taylor & Francis). Vol.17, No.1, pp.125-142, 2000.
- [62] J.V. Haxby, E.A. Hoffman, M.I. Gobbini: *The distributed human neural system for face perception*. Trends in Cognitive Sciences (Elsevier). Vol.4, No.6, pp.223-233, 2000.
- [63] V. Bruce, P.J.B. Hancock, and A.M. Burton: *Human face perception and identification*. In H. Wechsler, P.J. Phillips, V. Bruce, F.F. Soulié, and T.S. Huang, editors: *Face recognition: From theory to applications*. Vol.163, Part.1, Springer-Verlag, Berlin (Germany), pp.51-72, 1998.
- [64] R.K. Yin: *Looking at upside-down faces*. Journal of Experimental Psychology. Vol.81, No.1, pp.141-151, 1969.

- [65] J.C. Bartlett and J. Searcy: *Inversion and configuration of faces*. Cognitive Psychology (Elsevier). Vol.25, No.3, pp.281-316, 1993.
- [66] P. Thompson: *Margaret Thatcher: A new illusion*. Perception. Vol.9, No.4, pp.483-484, 1980.
- [67] S.E. Brennan: *Caricature generator: The dynamic exaggeration of faces by computer*. Leonardo (MIT Press). Vol.18, No.3, pp.170-178, 1985.
- [68] D.N. Perkins: *A definition of caricature and recognition*. Studies in the Anthropology of Visual Communication. Vol.2, pp.1-24, 1975.
- [69] V. Bruce, M. Burton, and N. Dench: *What's distinctive about a distinctive face ?*. The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology (Francis & Taylor). Vol.47, No.1, pp.119-141, 1994.
- [70] A.P. Ginsburg: *Visual information processing based on spatial filters constrained by biological data*. Doctoral Thesis. Reports of the Aerospace Medical Research Laboratory, AMLR-TR, Cambridge University (UK), pp.78-129, 1978.
- [71] L.D. Harmon: *The recognition of faces*. Scientific American. Vol.229, No.5, pp.71-82, 1973.
- [72] J. Sergent: *Microgenesis of face perception*. In H.D. Ellis, M.A. Jeeves, F. Newcombe, and A. Young, editors: *Aspects of face processing*. Vol.28, No.2, Springer-Verlag, Dordrecht (Netherlands), pp.17-33, 1986.
- [73] I. Biederman: *Recognition by components: A theory of human image understanding*. Psychological Review. Vol.94, No.2, pp.115-147, 1987.
- [74] H. Hill, P.G. Schyns, and S. Akamatsu: *Information and viewpoint dependence in face recognition*. Cognition (Elsevier). Vol.62, No.2, pp.201-222, 1997.
- [75] M.J. Tarr and H.H. Bulthoff: *Is human object recognition better described by geo-structural descriptions or by multiple views ? Comments on Biederman and Gerhardstein (1993)*. Journal of Experimental Psychology: Human Perception and Performance. Vol.21, No.6, pp.1494-1505, 1995.
- [76] D. Marr, editor: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt and Company, New York (USA), 1982.
- [77] H. Hill and V. Bruce: *Effects of lighting on the perception of facial surfaces*. Journal of Experimental Psychology: Human Perception and Performance. Vol.22, No.4, pp.986-1004, 1996.
- [78] A.J. O'Toole, D.A. Roark, and H. Abdi: *Recognizing moving faces: A psychological and neural synthesis*. Trends in Cognitive Sciences. Vol.6, No.6, pp.261-266, 2002.
- [79] B. Knight and A. Johnston: *The role of movement in face recognition*. Visual Cognition. Vol.4, No.3, pp.265-273, 1997.

- [80] X. Zhu and D. Ramanan: *Face detection, pose estimation, and landmark localisation in the wild*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp.2879-2886, Providence (USA), 2012.
- [81] A. Benzaoui, H. Bourouba, and A. Boukrouche: *System for automatic faces detection*. In Proceedings of the 3rd IEEE International Conference on Image Processing, Theory, Tools, and Applications (IPTA). pp.354-358, Istanbul (Turkey), 2012.
- [82] S.M.H. Anvar, W.Y. Yau, and E.K. Teoh: *Multiview face detection and registration requiring minimal manual intervention*. IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI). Vol.35, No.10, pp.2484-2497, 2013.
- [83] A. Martinez: *Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.24, No.6, pp.748-763, 2002.
- [84] J. Wu, S.C. Brubaker, M.D. Mullin, and J.M. Rehg: *Fast asymmetric learning for cascade face detection*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.30, No.3, pp.369-382, 2008.
- [85] W. Liao, D. Zeng, L. Zhou, S. Wang, and H. Zhong: *Wireless video surveillance system based on incremental learning face detection*. In proceedings of the 21st International Conference on Multimedia Modeling. Lecture Notes in Computer Science (Springer). Vol.8935, pp.118-127, Sydney (Australia), 2015.
- [86] C. Zhang and Z. Zhang: *Improving multiview face detection with multi-task deep convolutional neural networks*. In proceedings of the IEEE Winter Conference Applications of Computer Vision (WACV). pp.1036-1041, Colorado (USA), 2014.
- [87] A. Benzaoui and H.F. Merouani: *Hybrid system for robust faces detection*. Journal of Electronic Science and Technology (JEST). Vol.10, No.2, pp.167-172, 2012.
- [88] M. Kirby and L. Sirovich: *Application of the Karhunen-Loeve procedure for the characterization of human faces*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.12, No.1, pp.103-108, 1990.
- [89] M. Turk and P. Pentland: *Eigenfaces for recognition*. Journal of Cognitive Neuroscience. Vol.3, No.1, pp.71-86, 1991.
- [90] B. Scholkopf, A. Smola, and K.R. Muller: *Nonlinear component analysis as a kernel eigenvalue problem*. Neural Computation (MIT Press). Vol.10, No.5, pp.1299-1319, 1998.
- [91] L. Chengjun: *Gabor-based kernel PCA with fractional power polynomial models for face recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.26, No.5, pp.572-581, 2004.

- [92] D.D. Lee and H.S. Seung: *Learning the parts of objects by non-negative matrix factorization*. Nature. Vol.401, pp.788-791, 1999.
- [93] H. Yu and J. Yang: *A direct LDA algorithm for high-dimensional data - with application to face recognition*. Pattern Recognition (Elsevier). Vol.34, No.10, pp.2067-2070, 2001.
- [94] L. Juwei, K.N. Plataniotis, and A.N. Venetsanopoulos: *Face recognition using LDA-based algorithms*. IEEE Transactions on Neural Networks. Vol.14, No.1, pp.195-200, 2003.
- [95] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman: *Eigenfaces vs Fisherfaces: recognition using class specific linear projection*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.19, No.7, pp.711-720, 1997.
- [96] G. Baudat and F. Anouar: *Generalized discriminant analysis using a kernel approach*. Neural Computation (MIT Press). Vol.12, No.10, pp.2385-2404, 2000.
- [97] K. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf: *An introduction to kernel-based learning algorithms*. IEEE Transactions on Neural Networks. Vol.12, No.2, pp.181-201, 2001.
- [98] L. Juwei, K.N. Plataniotis, and A.N. Venetsanopoulos: *Face recognition using kernel direct discriminant analysis algorithms*. IEEE Transactions on Neural Networks. Vol.14, No.1, pp.117-126, 2003.
- [99] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Muller: *Fisher discriminant analysis with kernels*. In proceeding of the International IEEE Workshop on Neural Networks for Signal Processing. pp.41-48, Madison (USA), 1999.
- [100] S. Mika, G. Ratsch, B. Scholkopf, A. Smola, J. Weston, and K. Muller: *Invariant feature extraction and classification in kernel spaces*. Advances in Neural Information Processing (MIT Press). Vol.12, pp.526-532, 1999.
- [101] A. Martinez and A. Kak: *PCA versus LDA*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.23, No.2, pp.228-233, 2001.
- [102] A.K. Jain and B. Chandrasekaran: *Dimensionality and sample size considerations in pattern recognition practice*. In P.R. Krishnaiah and L.N. Kanal, editors: *Handbook of Statistics*. Vol.2, pp.835-855, Amsterdam (North-Holland), 1987.
- [103] S.J. Raudys and A.K. Jain: *Small sample size effects in statistical pattern recognition: recommendations for practitioners*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.13, No.3, pp.252-264, 1991.
- [104] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski: *Face recognition by independent component analysis*. IEEE Transactions on Neural Networks. Vol.13, No.6, pp.1450-1464, 2002.

- [105] L. Chengjun and H. Wechsler: *Independent component analysis of Gabor features for face recognition*. IEEE Transactions on Neural Networks. Vol.14, No.4, pp.919-928, 2003.
- [106] F.R. Bach and M.I. Jordan: *Kernel independent component analysis*. The Journal of Machine Learning Research. Vol.3, pp.1-48, 2002.
- [107] F.L. Bookstein: *Principal warps: Thin-plate splines and the decomposition of deformations*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.11, No.6, pp.567-585, 1989.
- [108] F.Y. Shih and C. Chuang: *Automatic extraction of head and face boundaries and facial features*. Information Sciences (Elsevier). Vol. 158, pp.117-130, 2004.
- [109] K. Sobottka and I. Pitas: *A fully automatic approach to facial feature detection and tracking*. In Proceedings of the 1st International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA). Lecture Notes in Computer Science (Springer). Vol.1206, pp.77-84, Crans-Montana (Switzerland), 1997.
- [110] M. Zobel, A. Gebhard, D. Paulus, J. Denzler, and H. Niemann: *Robust facial feature localization by coupled features*. In proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG). pp.2-7, Grenoble (France), 2000.
- [111] L. Wiskott, J.M. Fellous, and C.V. Der Malsburg: *Face recognition by elastic bunch graph matching*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.19, No.7, pp.775-779, 1997.
- [112] Z. Xue, S.Z. Li, and E.K. Teoh: *Bayesian shape model for facial feature extraction and recognition*. Pattern Recognition (Elsevier). Vol.36, No.12, pp.2819-2833, 2003.
- [113] B. Moghaddam, A. Pentland, and T. Starner: *View-based and modular eigenspaces for face recognition*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp.84-91, Washington (USA), 1994.
- [114] M. Tistarelli: *Active/space-variant object recognition*. Image and Vision Computing (Elsevier). Vol.13, No.3, pp.215-226, 1995.
- [115] M. Lades, J.C. Vorbuggen, J. Buhmann, J. Lange, C.V.D. Malsburg, R.P. Wurtz, and W. Konen: *Distortion invariant object recognition in the dynamic link architecture*. IEEE Transaction on Computers. Vol.42, No.3, pp.300-311, 1993.
- [116] L. Wiskott: *Phantom faces for face analysis*. Pattern Recognition (Elsevier). Vol.30, No.6, pp.837-846, 1997.
- [117] B. Duc, S. Fischer, and J. Bigun: *Face authentication with Gabor information on deformable graphs*. IEEE Transactions on Image Processing. Vol.8, No.4, pp.504-516, 1999.

- [118] C. Kotropoulos, A. Tefas, and I. Pitas: *Frontal face authentication using morphological elastic graph matching*. IEEE Transactions on Image Processing. Vol.9, No.4, pp.555-560, 2000.
- [119] P.T. Jackway and M. Deriche: *Scale-space properties of the multiscale morphological dilation-erosion*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.18, No.1, pp.38-51, 1996.
- [120] A. Tefas, C. Kotropoulos, and I. Pitas: *Face verification using elastic graph matching based on morphological signal decomposition*. Signal Processing (Elsevier). Vol.82, No.6, pp.833-851, 2002.
- [121] N. Kruger: *An algorithm for the learning of weights in discrimination functions using a priori constraints*. IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI). Vol.19, No.7, pp.764-768, 1997.
- [122] A. Tefas, C. Kotropoulos, and I. Pitas: *Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication*. IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI). Vol.23, No.7, pp.735-746, 2001.
- [123] K.I. Chang, K.W. Bowyer, and P.J. Flynn: *Face recognition using 2D and 3D facial data*. Workshop in Multimodal User Authentication. pp.25-32, California (USA), 2003.
- [124] K.I. Chang, K.W. Bowyer, and P.J. Flynn: *An evaluation of multimodal 2D+3D face biometrics*. IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI). Vol.27, No.4, pp.619-624, 2005.
- [125] C. Kotropoulos, A. Tefas, and I. Pitas: *Frontal face authentication using discriminating grids with morphological feature vectors*. IEEE Transactions on Multimedia. Vol.2, No.1 pp.14-26, 2000.
- [126] S. Arca, P. Campadelli, and R. Lanzarotti: *A face recognition system based on automatically determined facial fiducial points*. Pattern Recognition (Elsevier). Vol.39, No.3, pp.432-443, 2006.
- [127] Z. Chen, C. Liu, F. Chang, X. Han, and K. Wang: *Illumination processing in face recognition*. International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI). Vol.28, No.5, pp.1456011. 2014.
- [128] S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B.H. Romeny, J.B. Zimmerman, K. Zimmerman, K. Zuiderveld: *Adaptive histogram equalization and its variations*. Computer Vision, Graphics, and Image Processing. Vol.39, No.3, pp.355-368, 1987.
- [129] Y. Adini, Y. Moses, and S. Ullman: *Face recognition: The problem of compensating for changes in illumination direction*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.19, No.7, pp.721-732, 1997.
- [130] H.W. Jensen: *Digital face cloning*. In proceedings of the 30th International ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). Technical Report, University of California, pp.1-1, San Diego (USA), 2003.

- [131] L. Qing, S. Shan, and X. Chen: *Face relighting for face recognition under generic illumination*. In proceeding of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Vol.5, pp.733-736, Montreal (Canada), 2004.
- [132] R. Gross and V. Brajovic: *An image preprocessing algorithm for illumination invariant face recognition*. In Proceedings of the 4th International Conference on Audio- and Video-Based Biometric Person Authentication. Lecture Notes in Computer Science (Springer). Vol.2688, pp.10-18, Guildford (UK), 2003.
- [133] H. Wang, S.Z. Li, Y. Wang, and W. Zhang: *Illumination modeling and normalization for face recognition*. In Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG). pp.104-111, Nice (France), 2003.
- [134] L. Qing, S. Shan, and W. Gao: *Face recognition under varying lighting based on derivatives of log image*. In Proceedings of the 5th Chinese Conference on Biometric Recognition. Lecture Notes in Computer Science (Springer). Vol.3338, pp.196-204, Guangzhou (China), 2004.
- [135] Y. Weiss: *Deriving intrinsic images from image sequences*. In Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV). Vol.2, pp.68-75, Vancouver (Canada), 2001.
- [136] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek: *Overview of the face recognition grand challenge*. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Vol.1, pp.947-954, San Diego (USA), 2005.
- [137] C.L. Leveroni, M. Seidenberg, A.R. Mayer, L.A. Mead, J.R. Binder, S.M. Rao: *Neural systems underlying the recognition of familiar and newly learned faces*. The Journal of Neuroscience. Vol.20, No.2, pp.878-886, 2000.
- [138] J.R. Barr, K.W. Bowyer, P.J. Flynn, and S. Biswas: *Face recognition from video: A review*. International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI). Vol.26, No.5, pp.1266002, 2012.
- [139] H.K. Ekenel and R. Stiefelhagen: *Why is facial occlusion a challenging problem?*. In Proceedings of the 3rd IEEE/IAPR International Conference on Biometrics (ICB). Lecture Notes in Computer Science (Springer). Vol.5558, pp.299-308, Alghero (Italy), 2009.
- [140] R. He, W.S. Zheng, and B.G. Hu: *Maximum correntropy criterion for robust face recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.33, No.8, pp.1561-1576, 2011.
- [141] H. Jia and A.M. Martinez: *Face recognition with occlusions in the training and testing sets*. In Proceedings of the 8th IEEE International Conference on Face and Gesture Recognition (FG). pp.1-6, Amsterdam (Netherlands), 2008.

- [142] I. Naseem, R. Togneri, and M. Bennamoun: *Linear regression for face recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.32, No.11, pp.2106-2112, 2010.
- [143] A. Wanger, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma: *Toward a practical face recognition system: robust alignment and illumination by sparse representation*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.34, No.2, pp.372-386, 2012.
- [144] X. Wei, C.T. Li, and Y. Hu: *Robust face recognition under varying illumination and occlusion considering structured sparsity*. In Proceedings of the IEEE International Conference on Digital Image Computing Techniques and Applications (DICTA). pp.1-7, Fremantle (USA), 2012.
- [145] J. Wright, A.Y. Yang, A. Ganesh, and S.S. Sastry: *Robust face recognition via sparse representation*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.31, No.2, pp.210-227, 2009.
- [146] M. Yang and L. Zhang: *Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary*. In Proceedings of the 11th European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science (Springer). Vol.6316, pp.448-461, Heraklion (Greece), 2010.
- [147] D. Zhang, M. Yang, and X. Feng: *Sparse representation or collaborative representation: which helps face recognition?*. In Proceeding of the IEEE International Conference on Computer Vision (ICCV). pp.471-478, Barcelona (Spain), 2011.
- [148] X. Tan, S. Chen, Z.H. Zhou, and J. Liu: *Face recognition under occlusions and variant expressions with partial similarity*. IEEE Transactions on Information Forensics and Security. Vol.4, No.2, pp.217-230, 2009.
- [149] X. Wei, C.T. Li, and Y. Hu: *Face recognition with occlusion using dynamic image-to-class warping (DICW)*. In Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). pp.1-6, Shanghai (China), 2013.
- [150] X. Wei, C.T. Li, and Y. Hu: *Robust face recognition with occlusions in both reference and query images*. In Proceedings of the IEEE International Workshop on Biometrics and Forensics. pp.1-4, Lisbon (Portugal), 2013.
- [151] X. Wei and C.T. Li: *Fixation and saccade based face recognition from single image per person with various occlusions and expressions*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp.70-75, Portland (USA), 2013.
- [152] X. Tan, S. Chen, Z.H. Zhou, and F. Zhang: *Recognizing partially occluded, expresion variant faces from single training image per person with SOM and soft k-NN ensemble*. IEEE Transactions on Neural Networks. Vol.16, No.4, pp.875-886, 2005.
- [153] S. Liao, A.K. Jain, and S.Z. Li: *Partial face recognition: Alignment-free approach*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.35, No.5, pp.1193-1205, 2013.

- [154] J.C. Klontz and A.K. Jain: *A case study on unconstrained facial recognition using the Boston marathon bombings suspects*. Technical Report MSU-CSE. 2013.
- [155] B.F. Klare, Z. Li, and A.K. Jain: *Matching forensic sketches to mug shot photos*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.33, No.3, pp.639-646, 2011.
- [156] W. Zhang, X. Wang, and X. Tang: *Lighting and pose robust face sketch synthesis*. In Proceedings of the 11th European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science (Springer). Vol.6316, pp.420-433, Heraklion (Greece), 2010.
- [157] B.F. Klare and A.K. Jain: *Heterogeneous face recognition using kernel prototype similarities*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.35, No.6, pp.1410-1422, 2013.
- [158] X. Geng, Z.H. Zhou, K.S. Miles: *Automatic age estimation based on facial aging patterns*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.29, No.12, pp.2234-2240, 2007.
- [159] G. Guo, Y. Fu, C.R. Dyer, and T.S. Huang: *Image-based human age estimation by manifold learning and locally adjusted robust regression*. IEEE Transactions on Image Processing. Vol.17, No.7, pp.1178-1188, 2008.
- [160] A. Lanitis, C.J. Taylor, and T.F. Cootes: *Toward automatic simulation of aging effects on face images*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.24, No.4, pp.442-455, 2002.
- [161] Z. Li, U. Park, and A.K. Jain: *A discriminative model for age invariant face recognition*. IEEE Transactions on Information Forensics and Security. Vol.6, No.3, pp.1028-1037, 2011.
- [162] A.D. Dinkar and S.S. Sambyal: *Person identification in ethnic Indian goans using ear biometrics and neural networks*. Forensic Science International (Elsevier). Vol.223, No.(1-3), pp.373.e1-13, 2012.
- [163] H. Chen and B. Bhanu: *Human ear recognition in 3D*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.29, No.04, pp.718-737, 2007.
- [164] A. Benzaoui, A. Hadid, and A. Boukrouche: *Ear biometric recognition using local texture descriptors*. Journal of Electronic Imaging (JEI-SPIE). Vol.23, No.05, pp.053008.(01-12), 2014.
- [165] G.L. Streeter: *Developpement of the auricle in the human embryo*. Contributions to Embryology. Carnegie Institution of Washington (USA). Vol.14, pp.111-138, 1922.
- [166] K.K. Sulik: *Embryology of the ear*. In R.J. Gorlin, H.V. Toriello, and M.M. Cohen, editors: *Hereditary hearing loss and its syndromes*. Oxford Monographs on Medical Genetics. Oxford University Press - New York (USA). Vol.28, pp.22-42, 1995.

- [167] T.W. Sadler, editor: *Langman's medical embryology* (10th eds). Lippincott Williams & Wilkins, Philadelphia (USA). 2006.
- [168] J.E. Davis: *Surgical embryology*. In J.E. Davis, editor: *Aesthetic and reconstructive otoplasty*. Springer-Verlag, New York (USA). pp.93-125, 1987.
- [169] F. Wood-Jones and I.W. Chuan: *The development of the external ear*. *Journal of Anatomy*. Vol.68, pp.525-533, 1934.
- [170] C.S. Karmody and D.J. Annino: *Embryology and anomalies of the external ear*. *Facial Plastic Surgey Journal*. Vol.11, No.04, pp.251-256, 1995.
- [171] A.G Hunter and T. Yotsuyanagi: *The external ear: more attention to detail may aid syndrome diagnosis and contribute answers to embryological questions*. *American Journal of Medical Genetics* (Wiley). Vol.135, No.03, pp.237-250, 2005.
- [172] C. Park and T.S. Roh: *Congenital upper auricular detachment*. *Plastic and Reconstructive Surgey*. Vol.104, No.02, pp.488-490, 1999.
- [173] L. Meijerman: *Inter- and Intra individual variation in earprints*. Doctoral Thesis. Department of Anatomy and Embryology, Leiden University Medical Center (Netherlands), 2006.
- [174] B. Bhanu and H. Chen, editors: *Human ear recognition by computer*. *Advances in Computer Vision and Pattern Recognition Series*. Springer-Verlag, London (UK), 2008.
- [175] P. Viola and A.J. Jones: *Robust real-time face detection*. *International Journal of Computer Vision* (Springer). Vol.57, No.02, pp.137-154, 2004.
- [176] A. Iannarelli: *Ear Identification*. Paramount Publishing Company. Freemont California (USA), 1949.
- [177] M. Burge and W. Burger: *Ear Biometrics*. In A.K. Jain, R. Bolle, and S. Pankanti, editors: *Biometrics: Personal Identification in Networked Society*. Springer-Verlag, New York (USA), pp.273-285, 1999.
- [178] M. Burge and W. Burger: *Ear Biometrics in Computer Vision*. In *Proceedings of the 15th IEEE International Conference of Pattern Recognition (ICPR)*. Vol.02, pp.822-826, Barcelona (Spain), 2000.
- [179] D.J. Hurley, M.S. Nixon, and J.N. Carter: *Force field energy functionals for image feature extraction*. *Image and Vision Computing* (Elsevier). Vol.20, No.5-6, pp.311-317, 2002.
- [180] M. Abdel-Mottaleb and J. Zhou: *Human ear recognition from face profile images*. In *Proceedings of the International Conference on Biometrics (ICB)*. *Lecture Notes in Computer Science* (Springer). Vol.3832, pp.786-792, Hong Kong (China), 2006.
- [181] J. Dong and Z. Mu: *Multi-pose ear recognition based on force field transformation*. In *Proceedings of the 2nd IEEE International Symposium on Intelligent Information Technology Application (IITA)*. Vol.03, pp.771-775, Shanghai (China), 2008.

- [182] W. Liu, Y. Wang, S.Z. Li, and T. Tan: *Null space approach of fisher discriminant analysis for face recognition*. In Proceedings of the 8th European Conference on Computer Vision (ECCV) / Biometric Authentication Workshop (BIOAW). Lecture Notes in Computer Science (Springer). Vol.3087, pp.32-44, Prague (Czech Republic), 2004.
- [183] W. Liu, Y. Wang, S.Z. Li, and T. Tan: *Null space-based kernel fisher discriminant analysis for face recognition*. In Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition (FG). pp.369-374, Seoul (South Korea), 2004.
- [184] A. Kumar and C. Wu: *Automated human identification using ear imaging*. Pattern Recognition (Elsevier). Vol.45, No.03, pp.956-968, 2012.
- [185] X. Wang and W. Yuan: *Gabor wavelets and general discriminant analysis for ear recognition*. In Proceedings of the 8th IEEE World Congress on Intelligent Control and Automation (WCICA). pp.6305-6308, Jinan (China), 2010.
- [186] B. Arbab-Zavar and M.S. Nixon: *Robust log-gabor filter for ear biometrics*. In Proceedings of the 19th IEEE International Conference on Pattern Recognition (ICPR). pp.1-4, Florida (USA), 2008.
- [187] A.F. Abate, M. Nappi, D. Riccio, and S. Ricciardi: *Ear recognition by means of a rotation invariant descriptor*. In Proceedings of the 18th IEEE International Conference on Pattern Recognition (ICPR). Vol.04, pp.437-440, Hong Kong (China), 2006.
- [188] R. Fooprateepsiri and W. Kurutach: *Ear based personal identification approach forensic science tasks*. Chiang Mai Journal of Science. Vol.38, No.02, pp.166-175, 2011.
- [189] M. Petrou and A. Kadyrov: *The trace transform and its applications*. In Proceedings of the Nobless Workshop on Non-Linear Model Based Image Analysis (NMBIA). Springer-Verlag London. pp.207-214, Glasgow (Scotland), 1998.
- [190] A. Sana, P. Gupta, and R. Purkait: *Ear biometrics: a new approach*. In Proceedings of the 6th International Conference on Advances in Pattern Recognition (ICAPR). pp.1-5, Kolkata (India), 2007.
- [191] A. Basit and M. Shoaib: *A human ear recognition method using nonlinear curvelet feature subspace*. International Journal of Computer Mathematics (Taylor & Francis). Vol.91, No.03, pp.616-624, 2014.
- [192] E.J. Candes and D.L. Donoho: *Curvelets, multiresolution representation, and scaling laws*. In Proceedings of the SPIE Wavelet Applications in Signal and Image Processing VIII. Vol.4119, pp.1-12, San Diego (USA), 2000.
- [193] M. De Marsico, N. Michele, and D. Riccio: *HERO: human ear recognition against occlusions*. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp.178-183, San Francisco (USA), 2010.

- [194] X.Q. Wang, H.Y. Xia, and Z.L. Wang: *The research of ear identification based on improved algorithm of moment invariant*. In Proceedings of the 3rd IEEE International Conference on Information and Computing (ICIC). Vol.01, pp.58-60, Jiang Su (China), 2010.
- [195] X. Wang and W. Yuan: *Human ear recognition based on block segmentation*. In Proceedings of the IEEE International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC). pp.262-266, Zhangjiajie (China), 2009.
- [196] D.G. Lowe: *Object recognition from local scale-invariant features*. In Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV). Vol.02, pp.1150-1157, Kerkyra (Greece), 1999.
- [197] J.D. Bustard and M.S. Nixon: *Toward unconstrained ear recognition from two-dimensional images*. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans. Special Issue on Recent Advances in Biometrics. Vol.40, No.03, pp.486-494, 2010.
- [198] B. Arbab-Zavar, M.S. Nixon, and D.J. Hurley: *On model-based analysis of ear biometrics*. In Proceedings of the 1st IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS). pp.1-5, Crystal City (USA), 2007.
- [199] G.S. Badrinath and P. Gupta: *Feature level fused ear biometric system*. In Proceedings of the 7th IEEE International Conference on Advances in Pattern Recognition (ICAPR). pp.197-200, Kolkata (India), 2009.
- [200] D.R. Kisku, H. Mehrotra, P. Gupta, and J.K. Sing: *SIFT-based ear recognition by fusion of detected keypoints from color similarity slice regions*. In Proceedings of the IEEE International Conference on Advances in Computational Tools for Engineering Applications (ACTEA). pp.380-385, Beirut (Lebanon), 2009.
- [201] S. Prakash and P. Gupta: *An efficient ear recognition technique invariant to illumination and pose*. Telecommunication Systems (Springer). Vol.52, No.03, pp.1435-1448, 2013.
- [202] H. Bay, T. Tuytelaars, and L.V. Gool: *SURF: speeded up robust features*. In Proceedings of the 9th European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science (Springer). Vol.3951, pp.404-417, Graz (Austria), 2006.
- [203] M. Choras: *Perspective methods of human identification: ear biometrics*. Opto-Electronics Review (Springer).Vol.16, No.01, pp.85-96, 2008.
- [204] Z. Mu, L. Yuan, Z. Xu, D. Xi, and S. Qi: *Shape and structural feature based ear recognition*. In Proceedings of the 5th Chinese Conference on Biometric Recognition (SINOBIOMETRICS). Lecture Notes in Computer Science (Springer). Vol.3338, pp.663-670, Guangzhou (China), 2005.
- [205] M. Rahman, R. Islam, N.I. Bhuiyan, B. Ahmed, and A. Islam: *Person identification using ear biometrics*. International Journal of the Computer, the Internet and Management. Vol.15, No.02, pp.1-8, 2007.

- [206] F.N. Sibai, A. Nuaimi, A. Maamari, and R. Kuwair: *Ear recognition with feed-forward artificial neural network*. Neural Computing and Applications (Springer). Vol.23, No.05, pp.1265-1273, 2013.
- [207] Y. Guo and Z. Xu: *Ear recognition using a new local matching approach*. In Proceedings of the 15th IEEE International Conference on Image Processing (ICIP). pp.289-292, San Diego (USA), 2008.
- [208] Z.Q. Wang and X.D. Yan: *Multi-scale feature extraction algorithm of ear image*. In Proceedings of the International Conference on Electric Information and Control Engineering (ICEICE). pp.528-531, Wuhan (China), 2011.
- [209] E. Jeges and L. Mate: *Model-Based human ear localization and feature extraction*. International Journal of Intelligent Computing in Medical Sciences & Image Processing (Taylor & Francis). Vol.01, No.02, pp.101-112, 2007.
- [210] H. Liu and J. Yan: *Multi-view ear shape feature extraction and reconstruction*. In Proceedings of the 3rd IEEE International Conference on Signal-Image Technologies and Internet-Based Systems (SITIS). pp.652-658, Shanghai (China), 2007.
- [211] R. Mukundan, S.H. Ong, and P.A. Lee: *Image Analysis by Tchebichef Moments*. IEEE Transactions on Image Processing. Vol.10, No.09, pp.1357-1364, 2001.
- [212] L. Lu, Z. Xiaoxun, Z. Youdong, and J. Yunde: *Ear Recognition based on Statistical Shape Model*. In Proceedings of the 1st IEEE International Conference on Innovative Computing, Information and Control (ICICIC). Vol.03, pp.353-356, Beijing (China), 2006.
- [213] L. Yuan and Z.C. Mu: *Ear Recognition based on 2D Images*. In Proceedings of the 1st IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS). pp.1-5, Washington (USA), 2007.
- [214] B. Victor, K. Bowyer, and S. Sarkar: *An evaluation of face and ear biometrics*. In Proceedings of the 16th International Conference on Pattern Recognition (ICPR). Vol.01, pp.429-432, 2002.
- [215] K. Chang, K.W. Bowyer, S. Sarkar, and B. Victor: *Comparison and combination of ear and face images in appearance-based biometrics*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.25, No.09, pp.1160-1165, 2003.
- [216] M. Alaraj, J. Hou, and T. Fukami: *A neural network based human identification framework using ear images*. In Proceedings of the IEEE TENCON Conference. pp.1595-1600, Fukuoka (Japan), 2010.
- [217] H. Zhang and Z. Mu: *Compound structure classifier system for ear recognition*. In Proceedings of the IEEE International Conference on Automation and Logistics (ICAL). pp.2306-2309, Qingdao (China), 2008.
- [218] Z. Xie and Z. Mu: *Ear recognition using LLE and IDLLE algorithm*. In Proceedings of the 19th IEEE International Conference on Pattern Recognition (ICPR). pp.1-4, Tampa (USA), 2008.

- [219] L. Nanni and A. Lumini: *A multi-matcher for ear authentication*. Pattern Recognition Letters (Elsevier). Vol.28, No.16, pp.2219-2226, 2007.
- [220] T. Yuizono, Y. Wang, K. Satoh, and S. Nakayama: *Study on individual recognition for ear images by using genetic local search*. In Proceedings of the IEEE Congress on Evolutionary Computation (CEC). Vol.01, pp.237-242, Honolulu (USA), 2002.
- [221] M. Yaqubi, K. Faez, and S. Motamed: *Ear recognition using features inspired by visual cortex and support vector machine technique*. In Proceedings of the IEEE International Conference on Computer and Communication Engineering (ICCCCE). pp.533-537, Kuala Lumpur (Malaisie), 2008.
- [222] B. Moreno, A. Sanchez, J.F. Velez: *On the use of outer ear images for personal identification in security applications*. In Proceedings of the 33rd IEEE Annual International Carnahan Conference on Security Technology. pp.469-476, Madrid (Spain), 1999.
- [223] L. Gutierrez, P. Melin, and M. Lopez: *Modular neural network integrator for human recognition from ear images*. In Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN). pp.1-5, Barcelona (Spain), 2010.
- [224] I. Naseem, R. Togneri, and M. Bennamoun: *Sparse representation for ear biometrics*. In Proceedings of the 4th International Symposium on Visual Computing (ISVC). Lecture Notes in Computer Science (Springer). Vol.5359, pp.336-345, Las Vegas (USA), 2008.
- [225] B. Zhang, Z. Mu, H. Zeng, and S. Luo: *Robust ear recognition via nonnegative sparse representation of Gabor orientation information*. The Scientific World Journal (Hindawi). 2014 (Not yet assigned to a volume).
- [226] M. Tuceryan and A.K. Jain: *Texture analysis*. In C.H. Chen, L.F. Pau, and P.S.P. Wang, editors: *Handbook of Pattern Recognition and Computer Vision (2nd eds)*. World Scientific Publishing. pp.207-248, (Singapore),1998.
- [227] T. Ahonen, J. Matas, C. He, and M. Pietikäinen: *Rotation invariant image description with local binary pattern histogram fourier features*. In Proceedings of the 16th Scandinavian Conference (SCIA). Lecture Notes in Computer Science (Springer). Vol.5575, pp.61-70, Oslo (Norway), 2009.
- [228] E.R. Davies: *Introduction to texture analysis*. In M. Mirmehdi, X. Xie, and J. Suri, editors: *Handbook of texture analysis*. Imperial College Press. pp.1-31, London (UK), 2008.
- [229] M. Petrou and P.G. Sevilla: *Image Processing: Dealing with texture*. Wiley Online Library, 2006.
- [230] R.M. Haralick: *Statistical and structural approaches to texture*. Proceedings of the IEEE. Vol.67, No.5, pp.786-804, 1979.
- [231] L.V. Gool, P. Dewaele, and A. Oosterlinck: *Texture analysis anno 1983*. Computer Vision, Graphics, and Image Processing (Elsevier). Vol.29, No.3, pp.336-357, 1985.

- [232] H. Wechsler: *Texture analysis – A survey*. Signal Processing (Elsevier). Vol.2, No.3, pp.271-282, 1980.
- [233] B. Julesz: *Experiments in the visual perception of texture*. Scientific American. Vol.232, No.4, pp.34-43, 1975.
- [234] C.M. Wu and Y.C. Chen: *Statistical feature matrix for texture analysis*. Graphical Models and Image Processing (Elsevier). Vol.54, No.5, pp.407-419, 1992.
- [235] X. Xie and M. Mirmehdi: *A galaxy of texture features*. In M. Mirmehdi, X. Xie, and J. Suri, editors: *Handbook of texture analysis*. Imperial College Press. pp.375-406, London (UK), 2008.
- [236] T. Mäenpää and M. Pietikäinen: *Texture analysis with local binary patterns*. In C.H. Chen and P.S.P. Wang, editors: *Handbook of Pattern Recognition and Computer Vision (3rd eds)*. World Scientific Publishing. pp.197-216, (Singapore), 2005.
- [237] F. Tajeripour, M. Saberi, M. Rezaei, and S.F. Ershad: *Texture classification approach based on combination of random threshold vector technique and co-occurrence matrixes*. In Proceedings of the IEEE International Conference on Computer Science and Network Technology (ICCSNT). Vol.4, pp.2303-2306, Harbin (China), 2011.
- [238] Y. He and N. Sang: *Robust illumination invariant texture classification using gradient local binary patterns*. In Proceedings of the IEEE International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping. pp.1-6, Xiamen (China), 2011.
- [239] N. Sebe, M.S. Lew: *Texture features for content-based retrieval*. In M.S. Lew, editor: *Principles of Visual Information Retrieval*. Springer-Verlag, pp.51-85, London (UK), 2011.
- [240] O.A.B. Penatti, E. Valle, and R.S. Torres: *Comparative study of global color and texture descriptors for web image retrieval*. Journal of Visual Communication and Image Representation (Elsevier). Vol.23, No.2, pp.359-380, 2012.
- [241] R.O. Duda, P.E. Hart, and D.G. Stork: *Pattern classification (2nd eds)*. John Wiley & Sons, New York, 2001.
- [242] C.M. Bishop: *Pattern recognition and machine learning*. Springer-Verlag, New York, 2006.
- [243] D.J. Hand: *Classifier technology and the illusion of progress*. Statistical Science. Vol.21, No.1, pp.1-15, 2006.
- [244] T. Ojala, M. Pietikäinen, and D. Harwood: *A comparative study of texture measures with classification based on featured distribution*. Pattern Recognition (Elsevier). Vol.29, No.1, pp.51-59, 1996.
- [245] M. Varma and A. Zisserman: *A statistical approach to texture classification from single images*. International Journal of Computer Vision (Springer). Vol.62, No.1-2, pp.61-81, 2005.

- [246] E. Hayman, B. Caputo, M. Fritz, and J.O. Eklundh: *On the significance of real-world conditions for material classification*. In Proceedings of the 08th European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science (Springer). Vol.3024, pp.253-266, Prague (Czech Republic), 2004.
- [247] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid: *Local features and kernels for classification of texture and object categories: comprehensive study*. International Journal of Computer Vision (Springer). Vol.73, No.2, pp.213-238, 2007.
- [248] J.T. Todd: *The visual perception of 3D shape*. Trends in Cognitive Sciences. Vol.8, No.3, pp.115-121, 2004.
- [249] A. Lobay and D.A. Forsyth: *Shape from texture without boundaries*. International Journal of Computer Vision (Elsevier). Vol.67, No.1, pp.71-91, 2006.
- [250] J.S.D. Bonet and P. Viola: *A non-parametric multi-scale statistical model for natural images*. Advances in Neural Information Processing (MIT Press). Vol.10, pp.773-779, 1997.
- [251] J. Portilla and E.P. Simoncelli: *A parametric texture model based on joint statistics of complex wavelet coefficients*. International Journal of Computer Vision (Springer). Vol.40, No.1, pp.49-70, 2000.
- [252] A.A. Efros and W.T. Freeman: *Image quilting for texture synthesis and transfer*. In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). pp.341-346, New York (USA), 2001.
- [253] M.R. Turner: *Texture discrimination by Gabor functions*. Biological Cybernetics (Springer). Vol.55, No.2-3, pp.71-82, 1986.
- [254] A.C. Bovik, M. Clark, and W.S. Geisler: *Multichannel texture analysis using localized spatial filters*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.12, No.1, pp.55-73, 1990.
- [255] A.K. Jain and F. Farrokhnia: *Unsupervised texture segmentation using Gabor filters*. Pattern Recognition (Elsevier). Vol.24, No.12, pp.1167-1186, 1991.
- [256] B.S. Manjunah and W.Y. Ma: *Texture features for browsing and retrieval of image data*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.18, No.8, pp.837-842, 1996.
- [257] S.E. Grigorescu, N. Petkov, and P. Kruizinga: *Comparison of texture features based on Gabor filters*. IEEE Transactions on Image Processing. Vol.11, No.10, pp.1160-1167, 2002.
- [258] T. Leung and J. Malik: *Representing and recognizing the visual appearance of materials using three-dimensional textons*. International Journal of Computer Vision (Springer). Vol.43, No.1, pp.29-44, 2001.
- [259] S. Lazebnik, C. Schmid, and J. Ponce: *A sparse texture representation using local affine regions*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.27, No.8, pp.1265-1278, 2005.
- [260] C. Schmid, R. Mohr, and C. Bauckhage: *Evaluation of interest point detectors*. International Journal of Computer Vision (Springer). Vol.37, No.2, pp.151-172, 2000.

- [261] K. Mikolajczyk, Y. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L.V. Gool: *A comparison of affine region detectors*. International Journal of Computer Vision (Springer). Vol.65, No.1-2, pp.43-72, 2005.
- [262] E. Nowak, F. Jurie, and B. Triggs: *Sampling strategies for bag-of-features image classification*. In Proceedings of the 9th European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science (Springer). Vol.3954, pp.490-503, Graz (Austria), 2006.
- [263] T. Tuytelaars and C. Schmid: *Vector quantizing feature space with a regular lattice*. In Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV). pp.1-8, Rio de Janeiro (Brazil), 2007.
- [264] T. Ahonen, A. Hadid, and M. Pietikäinen: *Face description with local binary patterns: Application to face recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.28, No.12, pp.2037-2041, 2006.
- [265] J. Vogel and B. Schiele: *Semantic modeling of natural scenes for content-based image retrieval*. International Journal of Computer Vision (Springer). Vol.72, No.2, pp.133-157, 2007.
- [266] M. Mikolajczyk and C. Schmid: *A performance evaluation of local descriptors*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.27, No.10, pp.1615-1630, 2005.
- [267] N. Dalal and B. Triggs: *Histograms of oriented gradients for human detection*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Vol.1, pp.886-893, San Diego (USA), 2005.
- [268] M. Heikkilä, M. Pietikäinen, and C. Schmid: *Description of interest regions with local binary patterns*. Pattern Recognition (Elsevier). Vol.42, No.3, pp.425-436, 2009.
- [269] T. Ojala, M. Pietikäinen, and T. Mäenpää: *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.24, No.7, pp.971-987, 2002.
- [270] T. Mäenpää: *The local binary pattern approach to texture analysis – extensions and applications*. PhD Thesis, Acta Universitatis Ouluensis C.187, University of Oulu (Finland), 2003.
- [271] L. Houam: *Contribution à l'analyse de textures radiographiques osseuses pour le diagnostic précoce de l'ostéoporose*. Phd Thesis, University of Guelma (Algeria) + University of Orléans (France), 2013.
- [272] L. Houam, A. Hafiane, R. Jennane, A. Boukrouche, and A. Lespessailles: *Trabecular bone anisotropy characterisation using 1D local binary patterns*. In Proceedings of the 12th International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS). Lecture Notes in Computer Science (Springer). Vol.6474, pp.105-113, Sydney (Austria), 2010.

- [273] L. Houam, A. Hafiane, A. Boukrouche, E. Lespessailles, and R. Jennane: *One dimensional local binary pattern for bone texture characterization*. Pattern Analysis and Applications (Springer). Vol.17, No.1, pp.179-193, 2014.
- [274] H.B. Kim and R.H. Park: *Extraction of periodicity vectors from structural textures using projection information*. Pattern Recognition Letters (Elsevier). Vol.11, No.9, pp.625-630, 1990.
- [275] H.B. Kim and R.H. Park: *Extracting spatial arrangement of structural textures using projection information*. Pattern Recognition (Elsevier). Vol.25, No.3, pp.237-245, 1992.
- [276] I.Z. Yalniz and S. Aksoy: *Unsupervised detection and localization of structural textures using projection profiles*. Pattern Recognition (Elsevier). Vol.43, No.10, pp.3324-3337, 2010.
- [277] V. Ojansivu and J. Heikkila: *Blur insensitive texture classification using local phase quantization*. In Proceedings of the 3rd International Conference on Image and Signal Processing (ICIP). Lecture Notes in Computer Science (Springer). Vol.5099, pp.236-243, Cherbourg-Octeville (France), 2008.
- [278] J. Kannala and E. Rahtu: *BSIF: binarized statistical image features*. In Proceedings of the 21st International IEEE Conference on Pattern Recognition (ICPR). pp.1363-1366, Tsukuba (Japan), 2012.
- [279] A. Hyvarinen and E. Oja: *Independent component analysis: algorithms and applications*. Neural Networks. Vol.13, No.4, pp.411-430, 2000.
- [280] F. Samria and A. Harter: *Parameterisation of a stochastic model for human face identification*. In Proceedings of the IEEE Workshop on Applications of Computer Vision. pp.138-142, 1994.
- [281] A.M. Martinez and R. Benavente: *The AR face database*. CVC Technical Report 24, 1998.
- [282] IIT Delhi Ear Database: http://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Ear.htm
- [283] USTB Ear Database: <http://www1.ustb.edu.cn/resb/en/index.htm>
- [284] H.J. Oh, K.M. Lee, and S.U. Lee: *Occlusion invariant face recognition using selective local non-negative matrix factorization basis images*. Image and Vision Computing (Elsevier). Vol.26, No.11, pp.1515-1523, 2008.
- [285] R. Min, A. Hadid, and J. Dugelay: *Improving the recognition of faces occluded facial accessories*. In Proceedings of the 9th International IEEE Conference on Automatic Face & Gesture Recognition and Workshops (AFGR), pp.442-447, Santa Barbara (USA), 2011.
- [286] H.T. Nguyen, N.S. Vu, and A. Caplier: *How far we can improve micro features based face recognition Systems?*. In Proceedings of the 3rd International IEEE Conference on Image Processing, Theory, Tools, and Applications (IPTA). pp.350-353, Istanbul (Turkey), 2012.
- [287] H.T. Nguyen and A. Caplier: *Elliptical local binary patterns for face recognition*. In Proceedings of the International on Computer Vision with Local Binary Patterns Variants (with ECCV). Lecture Notes in Computer Science (Springer). Vol.7728, pp. 85-96, Daejeon (Korea), 2012.

- [288] X. Luan, B. Fang, L. Liu, W. Yang, and J. Qian: *Extracting sparse error for robust PCA for face recognition in the presence of varying illumination and occlusion*. Pattern Recognition (Elsevier). Vol.47, No.02, pp.495-508, 2014.
- [289] Y. Zhao, W.S. Chen, B. Pan, and B. Chen: *Supervised kernel construction for unsupervised PCA on face recognition*. In Proceedings of the 6th Chinese Conference on Pattern Recognition (CCPR). Lecture Notes in Computer Science (Springer). Vol.484, pp351-359, Changsha (China), 2014.
- [290] Y. Li, J. Liu, H. Lu, and S. Ma: *Learning robust face representation with classwise block-diagonal structure*. IEEE Transactions on Information Forensics and Security. Vol.9, No.12, pp.2051-2062, 2014.
- [291] T. Anam and A. Usman: *Person identification using ear recognition*. Telekomunica. Vol.10, No.02, pp.321-326, 2012.
- [292] T.S. Chan and A. Kumar: *Reliable ear identification using 2-D quadrature filters*. Pattern Recognition Letters (Elsevier). Vol.33, No.14, pp.1870-1881, 2012.
- [293] A. Kumar and T.S. Chan: *Robust ear identification using sparse representation of local texture descriptors*. Pattern Recognition (Elsevier). Vol.46, No.01, pp.73-85, 2013.
- [294] Mamta and H. Madasu: *Robust ear based authentication using local principal independent components*. Expert Systems with Applications (Elsevier). Vol.40, No.16, pp.6478-6490, 2013.
- [295] L. Jacob and G. Raju: *Ear recognition using texture features – A novel approach*. Advances in Signal Processing and Intelligent Recognition Systems (Springer). Vol.264, pp.1-12, 2014.

Annexe A

Notions fondamentales sur la classification, les k -plus proches voisins (k -NN) et les séparateurs à vaste marge (SVM)

L'annexe courante présente les outils que nous avons utilisés pour classer une identité inconnue à l'individu (la classe) correspondant(e).

A.1. Introduction à la classification

La classification, en général, se réfère au classement ou au groupement d'éléments de données dans des ensembles similaires. Cette information est souvent utile dans l'étape d'analyse pour n'importe quel système de traitement du signal ou de données. La classification a deux significations distinctes. Nous pouvons recevoir une série d'observations avec l'objectif d'établir l'existence des classes ou des groupes dans les données. Nous pouvons savoir avec certitude qu'il ya tant de classes, et l'objectif consiste à instaurer une règle selon laquelle nous pouvons classer une nouvelle observation dans l'une des classes existantes. Le premier type est connu comme l'apprentissage non supervisé et le second comme l'apprentissage supervisé.

La classification d'image est en général similaire à la classification des données, mais elle peut être différente en fonction de l'application dans laquelle elle est utilisée. La classification est souvent la dernière étape d'un processus général de reconnaissance des formes. Il s'agit généralement d'un tri d'objets dans une image ou plusieurs images dans des classes distinctes. Typiquement, l'image est segmentée ou traitée afin d'isoler les différents objets ou formes les uns des autres, et les différents objets ou images sont étiquetés. Une étape d'extraction de caractéristiques (attributs) réduit les données en mesurant certaines propriétés ou caractéristiques des objets ou images étiquetés. Ces attributs sont ensuite transmis à un classificateur qui évalue ces caractéristiques et prend une décision relative à la classe de chaque objet ou image. La qualité de l'image acquise dépend de la résolution, la sensibilité, la bande passante et du rapport signal sur bruit du système d'imagerie. Un prétraitement tel que le filtrage est souvent nécessaire. Les attributs extraits peuvent être transformés dans un

espace de caractéristiques alternative afin de produire de meilleures caractéristiques, avant d'être envoyés au classificateur.

L'objectif visé dans notre cas par la classification est de pouvoir distinguer entre N classes correspondantes et d'attribuer l'identité d'un individu inconnu à la classe correspondante (N: le nombre total des individus dans la base de données du système biométrique). Cette méthode est basée sur l'apprentissage suivi de la classification. La phase d'apprentissage correspond à l'extraction d'attributs caractéristiques à partir de l'image dans l'ensemble d'apprentissage. La classification est la phase au cours de laquelle sont utilisés les attributs précédemment extraits afin d'atteindre l'objectif initial.

A.1.1. Classification non-supervisée

Cette méthode de classification est aussi appelée "*classification automatique*", (Clustering en anglais) ou encore "*regroupement*". Aucune information à priori sur les classes n'est connue. Nous cherchons donc à regrouper les différents exemples en fonction de la valeur de leurs caractéristiques de manière à créer des classes homogènes. Nous supposons que nous disposons d'un ensemble d'objets que l'on note par $X = \{x_1, x_2, \dots, x_N\}$ caractérisé par un ensemble de caractéristiques "*D*", l'objectif du regroupement est de trouver les groupes auxquels appartient chaque objet "*x*" que nous notons par $C = \{c_1, c_2, \dots, c_N\}$. Ce qui revient à déterminer une fonction notée "*Ys*" qui associe à chaque élément de *X* un ou plusieurs éléments de *C*. Il faut pouvoir affecter une nouvelle observation à une classe. Les observations disponibles ne sont pas initialement identifiées comme appartenant à telle ou telle population. L'absence d'étiquette de classe est un lourd handicap qui n'est que très partiellement surmontable. Ce procédé nécessite généralement de fixer au préalable le nombre de classes désirées, que ce soit de manière empirique ou automatique. Parmi les méthodes non-supervisées les plus utilisées, nous citons deux types d'approches: les centres mobiles (*k*-means) et la classification hiérarchique.

A.1.2. Classification supervisée

Dans cette méthode de classification, nous disposons déjà d'exemples dont la classe est connue et étiquetée. Une information sur les données à traiter est disponible et utilisée pour entraîner le processus de classification, cela constitue la phase d'apprentissage du modèle. Cette information appelée ensemble d'apprentissage est généralement constituée d'un

ensemble d'individus {caractéristiques, classe associée}. Dans le cas de la classification des identités, l'ensemble d'apprentissage est constitué d'un ensemble d'identités "types". Chaque individu est donc composé du couple (caractéristiques de l'identité, classe associée). Cet ensemble est alors appris par un algorithme classique de classification supervisée parmi lesquels, nous citons: les k -plus proches voisins (k -NN), les réseaux de neurones, les séparateurs à vaste marge (SVM), etc. Une fois la phase d'apprentissage réalisée, l'algorithme de classification est alors utilisé afin de déterminer la classification d'un ensemble d'individus tests composé d'un grand nombre d'échantillons. Cette approche de classification d'identités supervisée reste un domaine de recherche très actif. Dans notre analyse des données biométriques, nous nous intéressons à cette approche pour attribuer une identité inconnue à son propre individu. Dans ce qui suit, nous allons décrire sommairement les classificateurs que nous avons utilisés dans le cadre de ce travail.

A.2. k -plus proches voisins (k -NN)

L'algorithme des k plus proches voisins (noté k -NN) fait partie des méthodes de classification les plus couramment utilisées. Il permet de traiter des nuages de points non linéairement séparables. Cette approche a l'avantage d'être à la fois facile et efficace. L'algorithme k -NN figure parmi les algorithmes simples d'apprentissage artificiel. Dans un contexte de classification d'une nouvelle observation x , l'idée fondatrice est de faire voter les plus proches voisins de cette observation. La classe de x est déterminée en fonction de la classe majoritaire parmi les k plus proches voisins de l'observation x . La méthode k -NN est donc une méthode basée sur le voisinage, non-paramétrique; ceci signifiant que l'algorithme permet de faire une classification sans faire d'hypothèse sur la fonction qui relie la variable dépendante aux variables indépendantes.

A.2.1. Méthode 1-NN

La méthode du plus proche voisin est une méthode non paramétrique où une nouvelle observation est classée dans la classe d'appartenance de l'observation de l'échantillon d'apprentissage qui lui est la plus proche, vis-à-vis des covariables utilisées. La détermination de leur similarité est basée sur des mesures de distance. Formellement, soit L l'ensemble de données à disposition ou échantillon d'apprentissage:

$$L = \{(y_i, x_i), i = 1, \dots, n_L\} \tag{A.1}$$

Où $y \in \{1, \dots, c\}$ désigne la classe de l'individu "i" et le vecteur $X_i = \{x_{i1}, \dots, x_{ip}\}$ représente les variables prédictives de l'individu "i". La détermination du plus proche voisin est basée sur une fonction distance notée $d(.,.)$.

La distance *euclidienne* ou *dissimilarité* entre deux individus caractérisés par "p" covariables est définie par:

$$d\left((x_1, x_2, \dots, x_p), (u_1, u_2, \dots, u_p)\right) = \sqrt{(x_1 - u_1)^2 + (x_2 - u_2)^2 + \dots + (x_p - u_p)^2} \quad (\text{A.2})$$

Ainsi, pour une nouvelle observation (y, x) le plus proche voisin $(y_{(1)}, x_{(1)})$ dans l'échantillon d'apprentissage est déterminé par:

$$d(x, x_{(1)}) = \min_i(d(x, x_{(i)})) \quad (\text{A.3})$$

Alors $\hat{y} = y_{(1)}$, la classe du plus proche voisin, est sélectionnée pour la prédiction de y . Les notations $x_{(j)}$ et $y_{(j)}$ représentent respectivement le $j^{\text{ème}}$ plus proche voisin de x et sa classe d'appartenance. La distance euclidienne est définie comme suit:

$$d(x_i, x_j) = [\sum_{s=1}^p (x_{is} - x_{js})^2]^{\frac{1}{2}} \quad (\text{A.4})$$

A.2.2. Méthode k-NN

La méthode des k plus proches voisins est une extension de l'idée précédente, qui est largement et communément utilisée en pratique. La plus proche observation n'est plus la seule observation utilisée pour la classification. Nous utilisons désormais les k plus proches observations. Ainsi la décision est en faveur de la classe majoritairement représentée par les k voisins. Soit k_A le nombre d'observations issues du groupe des plus proches voisins appartenant à la classe A .

$$\sum_{A=1}^c k_A = k \quad (\text{A.5})$$

Ainsi une nouvelle observation est prédite dans la classe l avec $l = \max_A(k_A)$.

Cela évite que la classe prédite ne soit déterminée seulement à partir d'une seule observation. La mise en œuvre de cette technique ne dépend que du paramètre k : pour $k = 1$, nous utilisons la méthode du seul plus proche voisin comme technique locale maximale, pour $k = l_n$, nous utilisons la classe majoritaire sur l'ensemble intégral des observations.

Le paramètre k doit être déterminé par l'utilisateur: $k \in N$. En classification binaire, il est utile de choisir k impair pour éviter les votes égalitaires. Le meilleur choix de k dépend du jeu de données. En général, les grandes valeurs de k réduisent l'effet du bruit sur la classification et donc le risque de sur-apprentissage, mais rendent les frontières entre classes moins distinctes. Il convient donc de faire un choix de compromis entre la variabilité associée à une faible valeur de k contre un sur-lissage (c'est à dire, gommage des détails) pour une forte valeur de k . Un bon k peut être sélectionné par diverses techniques heuristiques, par exemple, la validation-croisée.

La Figure A.1 présente cette méthode avec $k = 5$, pour des points image projetés dans un espace d'attributs de dimension $d = 2$ et pour un nombre de classe $N_C = 2$. Pour classer l'image-test, dont le point représentatif dans le sous-espace d'attributs est rouge, nous cherchons tout d'abord les 5 points prototypes les plus proches (au sens de la distance Euclidienne). Ces images sont celles présentées par des flèches avec le point représentatif de l'image à classer. Il y a 3 images appartenant à la classe C1 et 2 images appartenant à la classe C2. L'image-test est donc assignée à la classe C1.

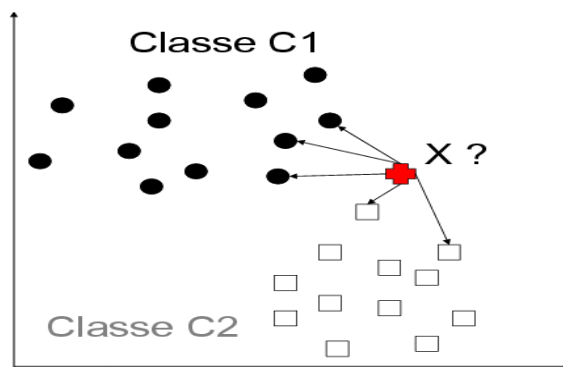


Figure A. 1 : Illustration de la méthode des k -plus proches voisins (k -NN) avec $k=5$.

Pour chaque image, nous obtenons un taux de classification qui est estimé par le rapport entre le plus proche voisin " k_i " de la classe correcte et le nombre total des plus proches voisins k :

$$P(C_i/x) = \frac{k_i}{k} \quad (\text{A.6})$$

$P(C_i/x)$ peut être considéré comme la probabilité a posteriori d'une instance x d'appartenir à la classe C_i selon la règle vote du k -NN. L'intérêt de ce type de classificateur est qu'il ne nécessite pas de réglage des paramètres préalables autres que le nombre des voisins " k ".

B.3. Séparateurs à vastes marges (SVM)

Cette méthode de classification de données a été proposée par *Vapnik* en 1995. Initialement, elle a été définie pour un problème de discrimination à deux classes (cas binaire). De manière simple, il s'agit de construire, dans l'espace des attributs un hyperplan séparant les points représentatifs des prototypes à chacune des deux classes. Cette approche, également appelée méthode des séparateurs à vastes marges, repose sur deux notions clés: la notion de marge maximale et la notion de noyau.

La marge est la distance séparant la frontière de séparation (l'hyperplan) et les prototypes les plus proches de cette frontière. Ces prototypes sont appelés *vecteurs supports*. Dans les SVM, la frontière de séparation est choisie comme étant celle qui maximise la marge (voir la Figure A.2).

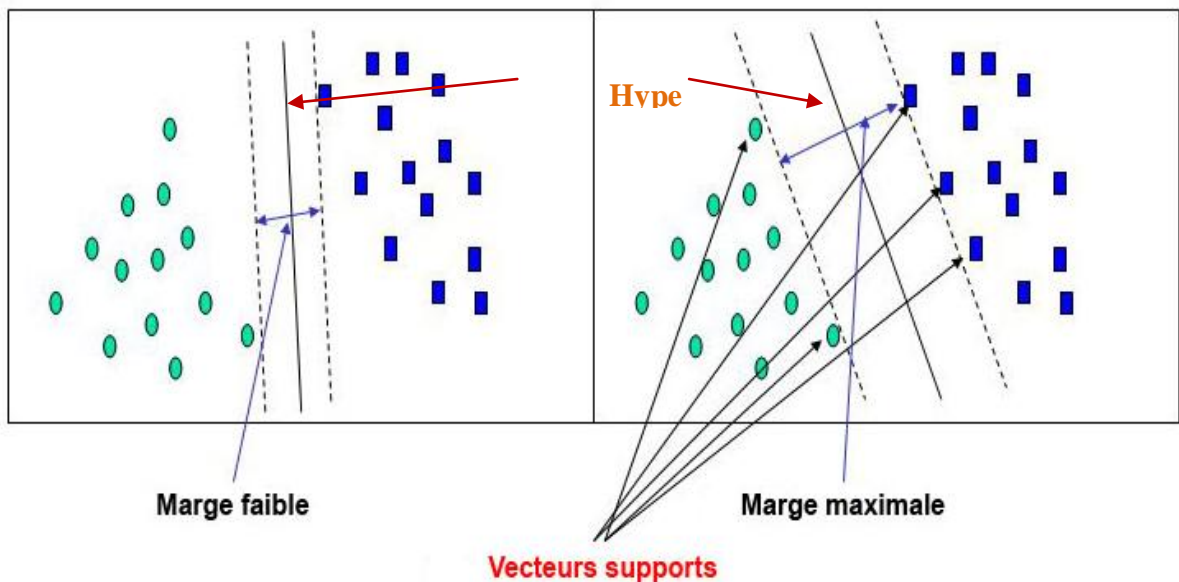


Figure A. 2 : Illustration de la marge maximale et les vecteurs supports.

Parmi les modèles des SVM, il y a les cas linéairement séparables (Figure A.2) et les cas non linéairement séparables. Les premiers sont les plus simples car ils permettent de trouver facilement le classificateur linéaire. Pour surmonter les inconvénients des cas non

linéairement séparables, l'idée des SVM est de changer l'espace des données. La transformation non linéaire des données peut permettre une séparation linéaire des exemples dans un nouvel espace.

Il s'agit d'une transformation d'un problème de séparation non linéaire dans l'espace de représentation en un problème de séparation linéaire dans un espace de re-description de plus grande dimension (voir figure A.3). Cette transformation non linéaire est réalisée via une fonction noyau. En pratique, quelques familles de fonctions noyau paramétrables sont connues et il revient à l'utilisateur de SVM d'effectuer des tests pour déterminer celle qui convient le mieux pour son application. Nous pouvons citer les exemples de noyaux suivants: Linéaire, Polynomiale, Gaussien, RBF (Radius Basis Function), Sigmoïde et Laplacien.

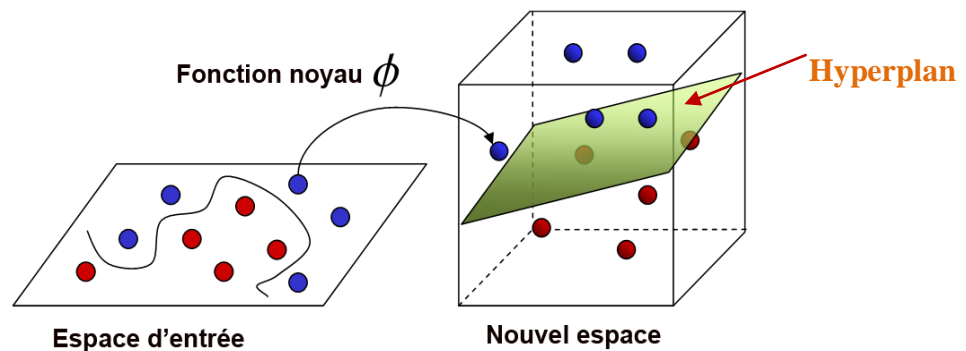


Figure A. 3 : Projection des données non linéairement séparables dans un espace à grande dimension.

Les séparateurs à vastes marges ont été initialement développés pour traiter des problèmes de classification binaire, mais ils peuvent être adaptés pour traiter les problèmes multi-classes. Il existe plusieurs stratégies pour faire la classification multi-classes. La première approche est de type "un contre tous": l'idée consiste simplement à transformer le problème à k classes en k classificateurs binaires et le classement est donné par le classificateur qui répond le mieux. La deuxième approche est de type "un contre un": cette fois le problème est transformée en $\frac{k \cdot (k-1)}{2}$ classificateurs binaires (chaque classe i étant en effet comparée à chaque classe j) et le classement est donné par le vote majoritaire ou un graphe acyclique de décision.

Notations

AR	Face Database of the Ohio State University
BSIF	Binarized Statistical Image Features
CCA	Canonical Correlation Analysis
CI	Personnes Correctement Identifiées
CMC	Cumulative Match Characteristics
DCT	Discrete Curvelet Transform
DNA	Deoxyribo Nucleic Acid
EBGM	Elastic Bunch Graph-Matching
EER	Equal Error Rate
EGM	Elastic Graph Matching
ELBP	Elliptical Local Binary Pattern
FA	False Acceptation
FAR	False-Acceptance Rate
FDCT	Fast Discrete Curvelet Transform
FERET	The Facial Recognition Technology Database
FLD	Fisher's Linear Discriminant Optimization
FNMR	False Non-Match Rate
FR	False-Rejection
FRGC	Face Recognition Grand Challenge Database
FRR	False-Rejection Rate
GDA	General Discriminant Analysis
GLCM	Gray Level Co-occurrence Matrix
GLOH	Gradient Location-Orientation Histogram
HOG	Histogram of Oriented Gradients
HSV	Hue Saturation Value
ICA	Independent Component Analysis
IHM	Interaction Homme Machine
IIT Delhi	Indian Institute of Technology Delhi
IITK	Indian Institute of Technology Kanpur
IR	Identification Rate
KFDA	Kernel Fisher Discriminant Analysis
K-ICA	Kernel Independent Component Analysis
k-NN	k-Nearest Neighbors
KPCA	Kernel Principal Component Analysis

LBP	Local Binary Pattern
LDA	Linear Discriminant Analysis
LLE	Locally Linear Embedding
LPQ	Local Phase Quantization
MEGM	Morphological Elastic Graph Matching
MRI	Magnetic Resonance Imaging
NKFDA	Null-Space Kernel Fisher Discriminant Analysis
NMF	Non-Negative Matrix Factorization
ORL	Olivetti Research Laboratory
PCA	Principal Component Analysis
PIFS	Partitioned Iterated Function System
PIMIS	Laboratory of Inverse Problems, Modeling, Information and Systems
PIN	Personal Identification Number
PSF	Point Spread Function
RBF	Radius Basis Function
RCBD	Representation with Class-wise Block-Diagonal
RF	Random Forests
RGB	The Color Space (Red, Green, Blue)
ROC	Receiver Operating Characteristic
ROI	Region of Interest
RPCA	Robust Principal Component Analysis
SFFS	Sequential Forward Floating Selection
SIFT	Scale Invariant Feature Transform
SK-PCA	Supervised Kernel for Principal Component Analysis
S-LNMF	Selective Local Nonnegative Matrix Factorization
SURF	Speeded Up Robust Features
SVD	Singular Value Decomposition
SVM	Support Vector Machines
SVR	Support Vector Regressors
UND	University of Notre Dame
USTB	University of Science and Technology Beijing
WPCA	Weighted Principal Component Analysis
1DLBP	One Dimensional Local Binary Pattern

Biometric Identification Using Local Texture Descriptors: Application to Face & Ear

Abstract:

Biometric systems are concerned with the automatic identification of people based on quantitative biological or physiological traits of the individual. In this research study, two biometric modalities are considered which are the face and the ear. Firstly, face recognition has several advantages compared to other technologies; it is natural, non-intrusive and easy to use. However, rapid and accurate identification of the face has a number of challenges and difficulties, especially in uncontrolled environments; in fact, four descriptors were tested and compared, namely: LBP (*Local Binary Pattern*), 1DLBP (*One Dimensional Local Binary Pattern*), LPQ (*Local Phase Quantization*) and BSIF (*Binarized Statistical Image Features*). Furthermore, the human ear is considered as a very recent modality in this field, it can be used as a counterpart to the face. Identification by ear remains until now in the course of exploration, its main advantage lies in its wealth of information and its stability over time.

In this thesis, we propose an approach of features extraction based mainly on local texture descriptors in order to recognize individuals from 2D intensity still images, of faces or ears under uncontrolled conditions. This approach has the ability to easily derive effective models of characteristics that combine in a single data vector the overall shape of the scanned object and the local texture of its appearance; it imitates, in a similar way, the capacity of the human being for object recognition, due to its particularity in codification of the components at low and high frequencies that contribute to the description of local and global features respectively.

Keywords: Biometrics, face, ear, identification, texture descriptors, LBP, 1DLBP, LPQ, BSIF.

التعريف البيومتري باستعمال واصفات النسيج المحلية: تطبيق على الوجه و الأذن

الملخص:

الأنظمة البيومترية توفر خاصية التحقق الآلي من الهوية استنادا إلى المقاييس البيولوجية للفرد. من خلال هذه الدراسة، اهتمنا بنوعين من النماذج البيومترية، المتمثلتان في الوجه والأذن. أولا، خاصية التعرف على الوجه لديها العديد من المزايا مقارنة بغيرها من التكنولوجيات، فهي طبيعية، غير تدخلية، وسهلة الاستخدام. غير أن التعرف السريع والدقيق على الوجه لا يزال لديه عدد من التحديات والصعوبات، وخاصة في البيانات غير الخاضعة للرقابة. ثانيا، تعتبر الأذن البشرية نموذجا حديثا جدا في هذا المجال، ويمكن استخدامها كقطعة مكمل للوجه. تحديد الهوية عن طريق الأذن لا يزال حتى الآن قيد الاستكشاف، ميزتها الرئيسية تكمن في غزارتها بالمعلومات واستقرارها على مر الزمن.

من خلال هذه الأطروحة، نقترح استخدام نموذجا جديدا لاستخراج المعلومات، مبنية أساسا على واصفات النسيج المحلية من أجل تحديد هوية الأفراد انطلاقا من صور ثنائية الأبعاد، متعلقة بالوجوه أو الأذنين، في ظل ظروف غير خاضعة للرقابة: أربعة واصفات جربت و قورنت، وهي: LBP (*Local Binary Pattern*), 1DLBP (*One Dimensional Local Binary Pattern*), LPQ (*Local Phase Quantization*), BSIF (*Binarized Statistical Image Features*). هذا النموذج لديه القدرة على استخلاص بسهولة خصائص فعالة من النماذج التي تجمع في ناقل بيانات واحد الشكل العام للجسم والنسيج المحلي من مظهره: انه يقلد، بطريقة مماثلة، قدرة الإنسان في التعرف على الأشياء، لخصوصيته في تدوين المكونات في الترددات المنخفضة والعالية التي تسهم في وصف الملامح المحلية و العامة على التوالي.

كلمات البحث: الأنظمة البيومترية، الوجه، الأذن، التعرف، واصفات النسيج، LBP، 1DLBP، LPQ، BSIF.